



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Projeto de um descritor para detecção de eventos em  
câmeras de videomonitoramento: Uma abordagem  
baseada na análise das características de altas  
frequências do fluxo ótico**

Ana Paula Gonçalves Soares de Almeida

Monografia apresentada como requisito parcial  
para conclusão do Curso de Engenharia da Computação

Orientador  
Prof. Dr. Flávio de Barros Vidal

Brasília  
2014

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Curso de Engenharia da Computação

Coordenador: Prof. Dr. Ricardo Zelenovsky

Banca examinadora composta por:

Prof. Dr. Flávio de Barros Vidal (Orientador) — CIC/UnB  
Prof. Dr. Alexandre Ricardo Soares Romariz — ENE/UnB  
Prof. Dr. Alexandre Zaghetto — CIC/UnB

### **CIP — Catalogação Internacional na Publicação**

de Almeida, Ana Paula Gonçalves Soares.

Projeto de um descritor para detecção de eventos em câmeras de videomonitoramento: Uma abordagem baseada na análise das características de altas frequências do fluxo ótico / Ana Paula Gonçalves Soares de Almeida. Brasília : UnB, 2014.

57 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2014.

1. câmera de segurança, 2. videomonitoramento, 3. tumulto,  
4. multidão, 5. fluxo ótico, 6. Transformada de Fourier, 7. detecção de evento

CDU 004

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



# Dedicatória

Dedico este trabalho aos meus pais, Maria Gorete Gonçalves Soares de Almeida e Rafael Soares de Almeida, pelo apoio e por tornarem esse sonho possível.



# Agradecimentos

Agradeço ao meu orientador, Prof. Dr. Flávio de Barros Vidal pela paciência, dedicação e confiança.

À família, pela compreensão, pela torcida e pelo apoio.

Aos colegas e amigos de curso: André Bauer, Rebeca Moura, Pedro Bispo e Ciro Miranda, por todos os momentos que passamos juntos e que tornaram essa etapa mais divertida.

# Resumo

Atualmente, o aumento do uso de câmeras de segurança em ambientes tanto abertos quanto fechados é inquestionável. Nossas ações diárias são monitoradas com maior frequência e isso traz uma confiança maior ao andarmos pelas ruas ou estabelecimentos quando não há uma grande movimentação.

Mas mesmo que exista vigilância constante, ainda há possibilidade de erros humanos ocorrerem e um tumulto passar totalmente despercebido, causando incômodos para a população, sendo esses eventos anormais emergenciais ou até fatais.

Este projeto visa propor uma abordagem para detectar eventos anormais com uma metodologia diferente das propostas atuais, utilizando uma combinação das técnicas de fluxo ótico e Transformada Discreta de Fourier deslocada e com dois métodos distintos de análise da distribuição espacial dos componentes de alta frequência do movimento.

Para trabalhos futuros, a criação de um sistema robusto de identificação de situações incomuns em cenários com grandes multidões é cogitada, além de tratar não apenas de câmeras fixas, mas também de vídeos de elevada variação de movimento.

**Palavras-chave:** câmera de segurança, videomonitoramento, tumulto, multidão, fluxo ótico, Transformada de Fourier, detecção de evento

# Abstract

Nowadays, the use of surveillance cameras in outside and inside environments is unquestionably higher. Our daily actions are monitored with more frequency and this brings a higher confidence when we walk in the streets or in buildings even when there is no movement.

Even if there is a constant vigilance, there is still the possibility of human errors occurring and an abnormal event would pass by, causing an uncomfortable moment for the population that could be dangerous or fatal.

This project proposes a different methodology to detect abnormal events in crowded environments, using a combination of optical flow techniques and shifted Discrete Fourier Transform with two distinct ways of analyzing the spacial distribution of the high-frequency motion components.

For further works, the development of a crowd event detection robust system is thought and besides treating fixed cameras, it will treat input videos with a high motion variation.

**Keywords:** surveillance camera, video monitoring, crowd, optical flow, Fourier transform, event detection

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>3</b>
2.1	Detecção de Eventos em Multidões . . . . .	3
2.2	Imagens em Movimento . . . . .	4
2.3	Fluxo Ótico . . . . .	5
2.3.1	Métodos . . . . .	7
2.4	Transformada de Fourier . . . . .	10
2.4.1	Impulso . . . . .	11
2.4.2	Convolução . . . . .	12
2.4.3	<i>Sampling</i> . . . . .	12
2.4.4	2-D . . . . .	13
2.4.5	Transformada Discreta de Fourier em 2-D . . . . .	13
2.4.6	Aplicações . . . . .	13
<b>3</b>	<b>Metodologia Proposta</b>	<b>15</b>
3.1	Sequência de Imagens de Entrada . . . . .	15
3.2	Fluxo Ótico . . . . .	17
3.3	Diferença entre Campo de Movimento Estimado . . . . .	17
3.4	Transformada Discreta de Fourier . . . . .	17
3.5	Deslocamento da Transformada de Fourier . . . . .	19
3.6	Extração das Componentes de Alta Frequência do Campo de Movimento Estimado . . . . .	21
3.6.1	Extração por Retângulos . . . . .	22
3.6.2	Extração por Arcos Retangulares . . . . .	22
<b>4</b>	<b>Resultados</b>	<b>26</b>
4.1	Base de Dados . . . . .	26
4.1.1	Vídeos Calmos . . . . .	27
4.1.2	Vídeos Agitados . . . . .	29

4.2	Testes Realizados . . . . .	31
4.3	Análise dos Resultados . . . . .	31
<b>5</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>43</b>
	<b>Referências</b>	<b>45</b>

# Lista de Figuras

2.1	Sequência de <i>frames</i> que compõem uma imagem em movimento, retirado de [21]. . . . .	4
2.2	Representação simplificada do vetor de fluxo ótico . . . . .	5
2.3	Vídeo original. . . . .	6
2.4	Campo de movimento da Figura 2.3. . . . .	6
2.5	Componente horizontal de velocidade da Figura 2.3. . . . .	9
2.6	Componente vertical de velocidade da Figura 2.3. . . . .	9
2.7	Derivadas parciais aproximadas da imagem, retirado de [5] . . . . .	10
2.8	Uma função periódica representada pela soma de senos e cossenos, retirada de [11]. . . . .	11
2.9	Exemplos de filtros passa-alta . . . . .	14
3.1	Fluxograma metodológico. . . . .	16
3.2	Vídeo original. . . . .	17
3.3	Componente horizontal do campo de movimento da Figura 3.2. . . . .	18
3.4	Componente vertical do campo de movimento da Figura 3.2. . . . .	18
3.5	Magnitudes da componente horizontal do fluxo ótico da Figura 3.2. . . . .	19
3.6	Magnitudes da componente vertical do fluxo ótico da Figura 3.2. . . . .	20
3.7	Deslocamento para o centro da componente horizontal do fluxo ótico da Figura 3.2. . . . .	20
3.8	Deslocamento para o centro da componente vertical do fluxo ótico da Figura 3.2. . . . .	21
3.9	Análise por retângulos. . . . .	22
3.10	Análise do componente vertical com cinco retângulos da Figura 3.2. . . . .	23
3.11	Análise por arcos retangulares. . . . .	23
3.12	Análise com três arcos retangulares da Figura 3.2. . . . .	24
3.13	Análise com três arcos retangulares da Figura 3.2. . . . .	24
3.14	Análise com três arcos retangulares da Figura 3.2. . . . .	25

4.1	À esquerda, pessoas caminhando pela Abbey Road. À direita, turistas em Wall Street. . . . .	27
4.2	À esquerda, pedestres em Dublin. À direita, imagem da Times Square. . .	27
4.3	À esquerda, transeuntes na Times Square. À direita, rua na Itália. . . . .	27
4.4	À esquerda, movimentação normal na Times Square. À direita, pedestres na Quinta Avenida. . . . .	28
4.5	À esquerda, praia do Rio de Janeiro. À direita, pessoas atravessando um cruzamento em Shibuya. . . . .	28
4.6	À esquerda, tumulto em show. À direita, ocorrência de uma briga em um show. . . . .	29
4.7	À esquerda, briga em um supermercado. À direita, ocorrência de uma briga em um show. . . . .	29
4.8	À esquerda, briga entre manifestantes. À direita, briga em um saguão. . .	29
4.9	À esquerda, pedestres em fuga. À direita, ocorrência de uma briga em um bar. . . . .	30
4.10	À esquerda, tumulto em um cassino. À direita, briga em uma loja de conveniência. . . . .	30
4.11	Gráficos para três cortes retangulares . . . . .	33
4.12	Gráficos para cinco cortes retangulares . . . . .	34
4.13	Gráficos para sete cortes retangulares . . . . .	35
4.14	Gráficos para onze cortes retangulares . . . . .	36
4.15	Gráficos para três cortes por arcos retangulares . . . . .	37
4.16	Gráficos para cinco cortes por arcos retangulares . . . . .	38
4.17	Gráficos para sete cortes por arcos retangulares . . . . .	39
4.18	Gráficos para onze cortes por arcos retangulares . . . . .	40

# Lista de Tabelas

4.1	Descrição Completa dos Vídeos Calmos Utilizados. . . . .	28
4.2	Descrição Completa dos Vídeos Agitados Utilizados. . . . .	30
4.3	Coordenadas $(x, y)$ para os vídeos calmos com extração retangular. . . . .	41
4.4	Coordenadas $(x, y)$ para os vídeos agitados com extração retangular. . . . .	41
4.5	Coordenadas $(x, y)$ para os vídeos calmos com extração por arcos retangulares. . . . .	42
4.6	Coordenadas $(x, y)$ para os vídeos agitados com extração por arcos retangulares. . . . .	42



# Capítulo 1

## Introdução

Atualmente, a maioria das situações alarmantes ocorrem em ambientes dinâmicos com muitas pessoas, por exemplo: brigas em estádios de futebol, confusões em shows, crimes em festas. Uma maneira encontrada para poder controlar a sensação constante de insegurança da população é instalação câmeras de videomonitoramento em locais públicos ou de grande movimento, como praças, estações de metrô, avenidas movimentadas e em locais fechados, como em estádios, centros comerciais.

O controle dessas câmeras geralmente não é automatizado, há uma dependência humana para detectar qualquer ação anormal que esteja ocorrendo. Existem centenas dessas câmeras em um mesmo ambiente de controle, logo, a equipe que monitora esses equipamentos pode facilmente inutilizar o objetivo principal das câmeras de segurança, a proteção da integridade da população.

Uma solução seria automatizar essa análise de multidões para fazer a detecção de ocorrências anormais e emitir alertas se faz necessária para o auxílio das autoridades responsáveis[15]. Para tal, as técnicas para desenvolver essa aplicação devem ser confiáveis e eficientes.

O interesse na área de comportamento de multidões é crescente, por possuir uma grande complexidade e por sua imensa possibilidade de abstração[14]. Existem diversos obstáculos para fazer a avaliação do comportamento, como oclusão, mudanças de iluminação, ou qualquer outro impedimento que possa influenciar o processo de detecção. Além disso, o fato de haver muitas pessoas em um mesmo ambiente dificulta a análise desses eventos e ocorrerão diversas ambiguidades no rastreamento.

Este projeto propõe o desenvolvimento inicial de um sistema robusto capaz de detectar situações anormais em multidões utilizando as técnicas de fluxo ótico e as informações das altas frequências da Transformada de Fourier, baseando-se na seguinte hipótese: É possível extrair informações de comportamento na imagem a partir das componentes de alta frequência da Transformada de Fourier do fluxo ótico diferencial.

A primeira etapa desenvolvida para a construção desse sistema (e a abordada nesse projeto), é a criação de um descritor. O descritor proposto visa trazer uma nova abordagem de detecção de eventos em ambientes com um grande número de pessoas. Com ele, é possível fazer uma distinção do que é considerado um cenário calmo ou um cenário conturbado, passando assim para a fase de classificação, que será um dos objetivos futuros do projeto.

A divisão do trabalho consiste em: o Capítulo 2 traz os principais conceitos teóricos utilizados para a produção do trabalho, além de conter alguns trabalhos relacionados a esse projeto. A metodologia proposta se encontra no Capítulo 3, os resultados obtidos estão no Capítulo 4. Os trabalhos futuros e conclusão estão no Capítulo 5.

# Capítulo 2

## Revisão Bibliográfica

Para a construção do descritor, é necessário abordar alguns conceitos técnicos e trabalhos correlatos. A separação do Capítulo 2 ocorre da seguinte maneira: a Seção 2.1 trata da detecção de eventos em multidões, a Seção 2.2 explica o conceito de imagens em movimento, a Seção 2.3 trata dos aspectos teóricos do fluxo ótico e na Seção 2.4 se encontra a Transformada Discreta de Fourier para imagens.

### 2.1 Detecção de Eventos em Multidões

Com a maior utilização das câmeras de monitoramento, deve-se considerar uma maneira mais eficaz de detecção de agitações, para evitar que fatores externos que frequentemente influenciam a capacidade cognitiva humana possam comprometer o gerenciamento das câmeras. Um método automatizado para remover a dependência total de um indivíduo é de extrema relevância.

A detecção de eventos anormais em multidões é uma questão de importância na área de segurança pública [19, 15]. Principalmente em espaços públicos, como estações de metrô, bancos, praças, entre outros. Segundo [18], a análise de eventos em multidões é um ponto crítico de pesquisa. Os maiores problemas nessa área são as dificuldades de detectar e rastrear identidades independentes, pois a maioria dos objetos em movimento sofrem oclusões severas. E também a complexidade em descrever as relações espaciais dos objetos.

Em [10], quando há uma dificuldade em detectar e rastrear precisamente um único objeto em uma cena de multidão, as soluções mais comuns dependem diretamente de uma análise em baixo nível das características do vídeo, através da detecção de movimento em região.

Alguns trabalhos recentes vêm apresentando maneiras de capturar esses momentos de tumulto de forma automatizada. Essas técnicas detectam o tempo exato em que algum

evento, como ações violentas, ocorreu. Por exemplo, em [22] é proposto um modelo de classificação Bag-of-Words (*BoW*). Já em [24], há uma fusão entre *BoW* e o algoritmo Motion SIFT (*MoSIFT*) para aperfeiçoar os resultados anteriores de [22].

Em [6], um novo modelo de característica de movimento, chamado Motion Co-Occurrence Feature (*MCF*), e um modelo baseado em energia é proposto, utilizando informações da velocidade de movimento estimada do alvo através do fluxo ótico e uma abordagem entrópica para características de desordem.

## 2.2 Imagens em Movimento

Imagens em movimento, ou vídeos, são uma sequência de imagens fixas (*frames*). Para que as imagens fixas deem uma impressão de movimento, estas devem possuir um intervalo determinado de continuidade. Esse intervalo é denominado taxa de *frame*[17] ou *frames per second*, frames por segundo (FPS).

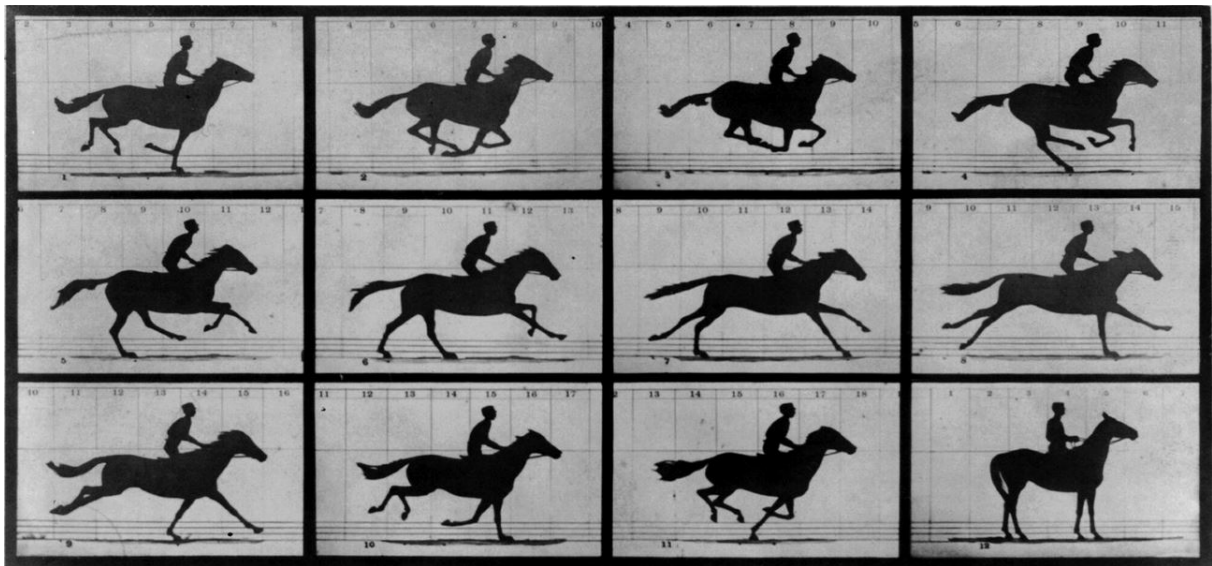


Figura 2.1: Sequência de *frames* que compõem uma imagem em movimento, retirado de [21].

O processamento de vídeos é feito *frame a frame*, dificultando, por vezes, um processamento em tempo real, devido à grande quantidade de dados que o programa deve lidar.

Pode-se afirmar que os vídeos possuem as mesmas características das imagens, por exemplo: canais, tamanho da imagem – altura e largura – e espaços de cor. Isso quer dizer que todo processamento que pode ser feito em uma imagem fixa pode também ser feito em um vídeo[3].

Uma imagem é uma representação em duas dimensões do mundo real, que possui três dimensões. Com uma sequência de imagens é possível extrair informações espaço-temporais que permitem uma reconstrução de objetos em um vídeo através de suas dimensões e posicionamento para com a câmera.

## 2.3 Fluxo Ótico

O fluxo ótico representa a movimentação aparente do padrão de brilho de um frame anterior para o frame atual[9]. O fluxo ótico é representado por um vetor, que descreve o movimento, que tem seu início no ponto inicial do *frame* anterior e sua direção definida para onde esse ponto se moveu, como mostra a Figura 2.2. Ele faz uma aproximação do campo de movimento da imagem representando o movimento aparente do padrão brilho dela no plano da imagem, em Figura 2.4. A Figura 2.4 é o campo de movimento aparente do fluxo ótico, mas por se tratar de uma imagem grande não é possível visualizar os vetores de movimento de fato. Esses se encontram na cor azul na imagem.

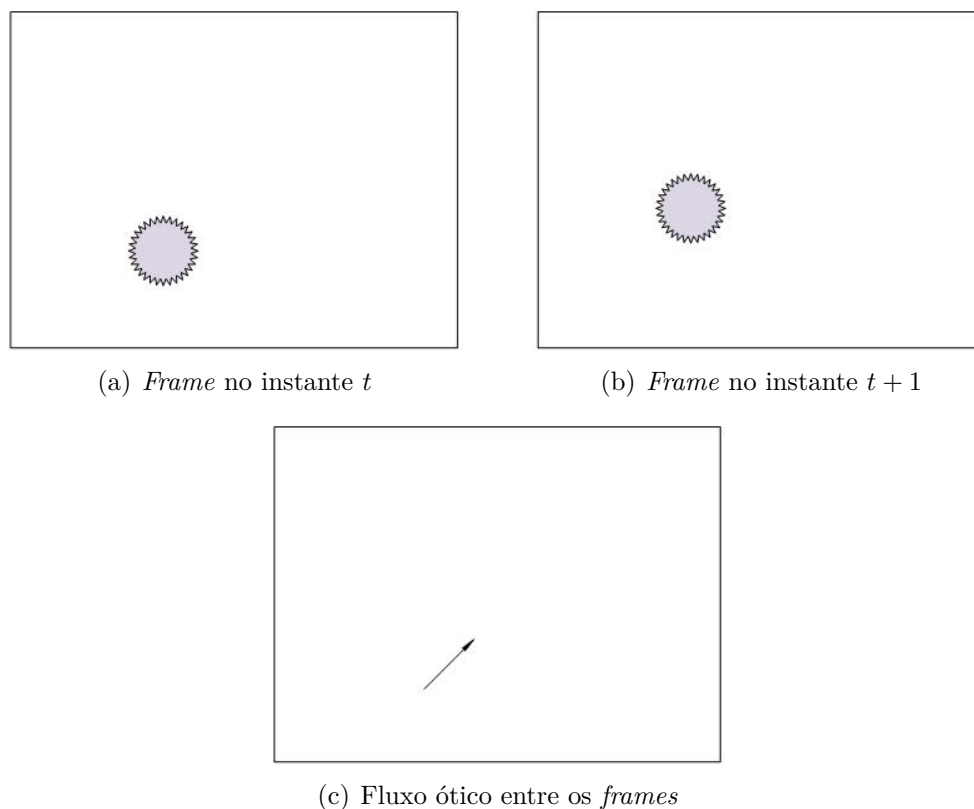


Figura 2.2: Representação simplificada do vetor de fluxo ótico



Figura 2.3: Vídeo original.

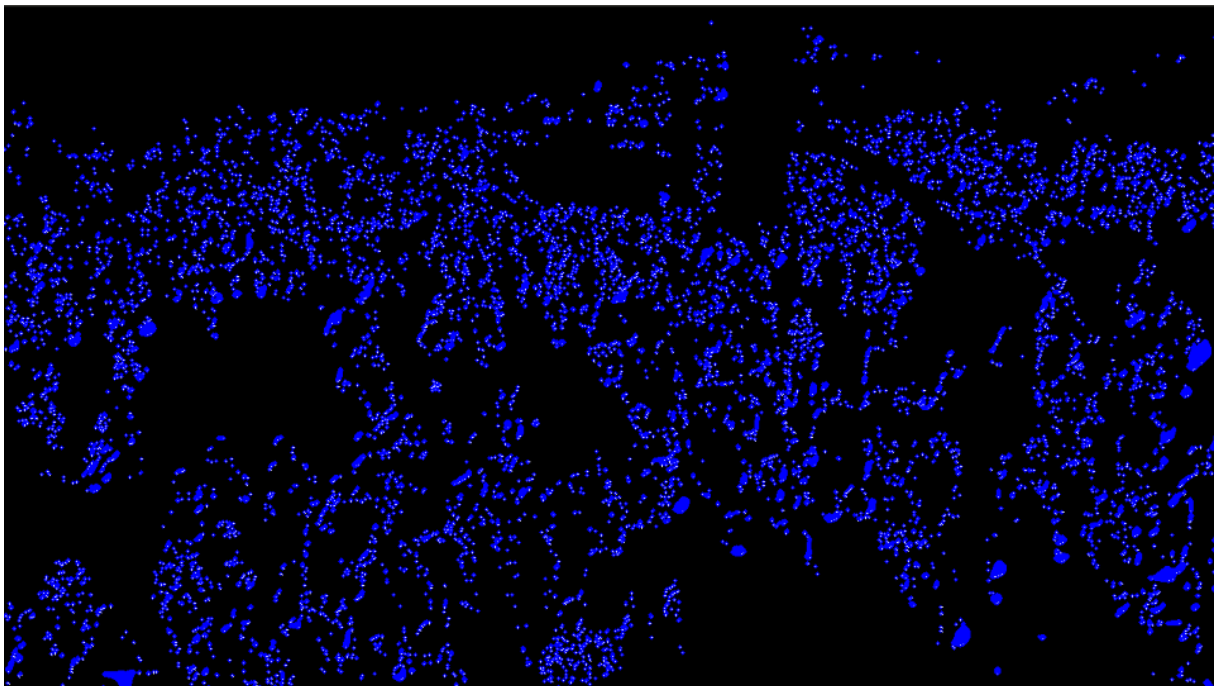


Figura 2.4: Campo de movimento da Figura 2.3.

Para determinar o fluxo ótico, deve-se considerar dois aspectos: o nível de precisão adquirido para intensidade e direção do movimento e a carga mínima necessária para que

o fluxo ótico possa desempenhar seu papel sob circunstâncias mínimas de exatidão.

### 2.3.1 Métodos

O fluxo ótico pode ser obtido por meio de três métodos distintos, sendo eles: *matching* baseado em região, métodos baseados em energia e fase e métodos diferenciais[2].

#### **Matching Baseado em Região**

Para a técnica de *matching* baseado em região, a velocidade  $v$  é o deslocamento  $d = (d_x, d_y)$ , onde  $d$  representa a similaridade, que leva ao melhor resultado entre regiões de imagens. Para encontrar a melhor correspondência, é preciso maximizar o valor de  $d$  e uma das maneiras de fazê-lo é utilizar a soma das diferenças quadráticas (SSD), mostrado na Equação 2.1[2]. A SSD é a maneira mais comum de medir as similaridades entre os blocos.

$$SSD_{1,2}(x; d) = \sum_{j=-n}^n \sum_{i=-n}^n W(i, j) [I_1(x+(i, j)) - I_2(x+d+(i, j))]^2 = W(x) * [I_1(x) - I_2(x+d)]^2 \quad (2.1)$$

Onde  $W$  representa uma função de custo  $2-d$  discreta.

Apesar da Equação 2.1 ter uma relação com as técnicas diferenciais (na seção 2.3.1), por se tratar de uma aproximação de primeira ordem da derivada temporal de  $I(x, t)$ , a técnica de *matching* baseado em região é utilizada quando há problemas de ruídos ou serrilhamentos (*aliasing*) na imagem, sendo, nesses casos, o resultado das técnicas diferenciais pouco precisos[1].

#### **Método Baseado em Energia e Fase**

Este método, também chamado de baseado em frequência, utiliza a energia da saída de filtros de velocidade para fazer o cálculo do fluxo ótico. considera-se também que a velocidade do fluxo ótico é calculada pelo comportamento da fase na saída de filtros passa-baixa[8].

O modelo de fluxo ótico foi implementado através da observação de como o movimento era percebido na neurofisiologia e na psicofísica e que o diferencia dos demais é que esse método utiliza um grande número de amostras de *frames* em conjunto no tempo[12].

Uma das primeiras técnicas de fase, Fleet e Jepson[8], capta as respostas espaço-temporais de filtros identicamente sintonizados. O componente de velocidade é dado por uma derivada de primeira ordem de uma superfície de fase constante.

Em [23], um filtro espaço-temporal é aplicado em mapas de bordas binárias para rastrear essas bordas. Para suavizar as bordas, um filtro Gaussiano é utilizado e sua saída gera um nível de contornos que são depois rastreados usando métodos diferenciais.

### Técnicas Diferenciais

As técnicas diferenciais calculam a velocidade pelas derivadas espaço-temporais da intensidade da imagem ou com o uso de filtros tanto passa-baixa quanto passa-alta. Essa velocidade pode ser calculada por meio de equações das derivadas de primeira ou segunda ordem do padrão de brilho do movimento[7].

Dentre as técnicas diferenciais, a mais conhecida é a de Horn e Schunck[13]. Este utiliza derivadas espaço-temporais da intensidade do brilho da imagem. De acordo com Horn e Schunck, o fluxo ótico não pode ser calculado em um ponto da imagem independentemente dos pontos vizinhos sem que haja limitações adicionais. Isso ocorre porque o campo de velocidade em cada ponto da imagem tem dois componentes, enquanto a mudança no brilho naquele ponto possui apenas uma restrição, devido ao movimento.

Assume-se que a velocidade aparente dos padrões de brilho podem ser diretamente identificados com o movimento das superfícies na cena. Isso implica, de acordo com a superfície do objeto que se move, que não há, ou existe pouca, variação de brilho. Um exemplo de ocorrência seria com objetos de simetria radial, pouco contraste global e alto nível de reflectância especular. Assume-se também que a iluminação incidente é constante.

Denota-se  $I(x, y, t)$  como o brilho da imagem  $t$  no ponto da imagem  $(x, y)$ . Durante o movimento, supõe-se que o brilho de um ponto particular é constante, isso significa que

$$\frac{dI(x, y, t)}{dt} = 0 \quad (2.2)$$

Expandindo e reescrevendo a Equação 2.2

$$I_x u + I_y v + I_t = 0 \quad (2.3)$$

Onde  $I_x$ ,  $I_y$  e  $I_t$  representam as derivadas parciais de brilho em  $x$ ,  $y$  e  $t$ , respectivamente;  $u$  e  $v$  são os componentes de velocidade  $x$  e  $y$ .

Considerando isso, o padrão de brilho pode se mover independentemente do resto da cena e há possibilidade de recuperação da informação da velocidade, como mostrado nas Figuras 2.5 a 2.6, respectivamente.

As derivadas parciais do brilho da imagem são aproximadas de um conjunto discreto de medidas dos brilhos da imagem. Para evitar problemas causados por valores zerados para as derivadas nas direções espaço-temporais, o ponto de interesse está localizado no centro de um cubo formado por oito medidas, apresentado na Figura 2.7.



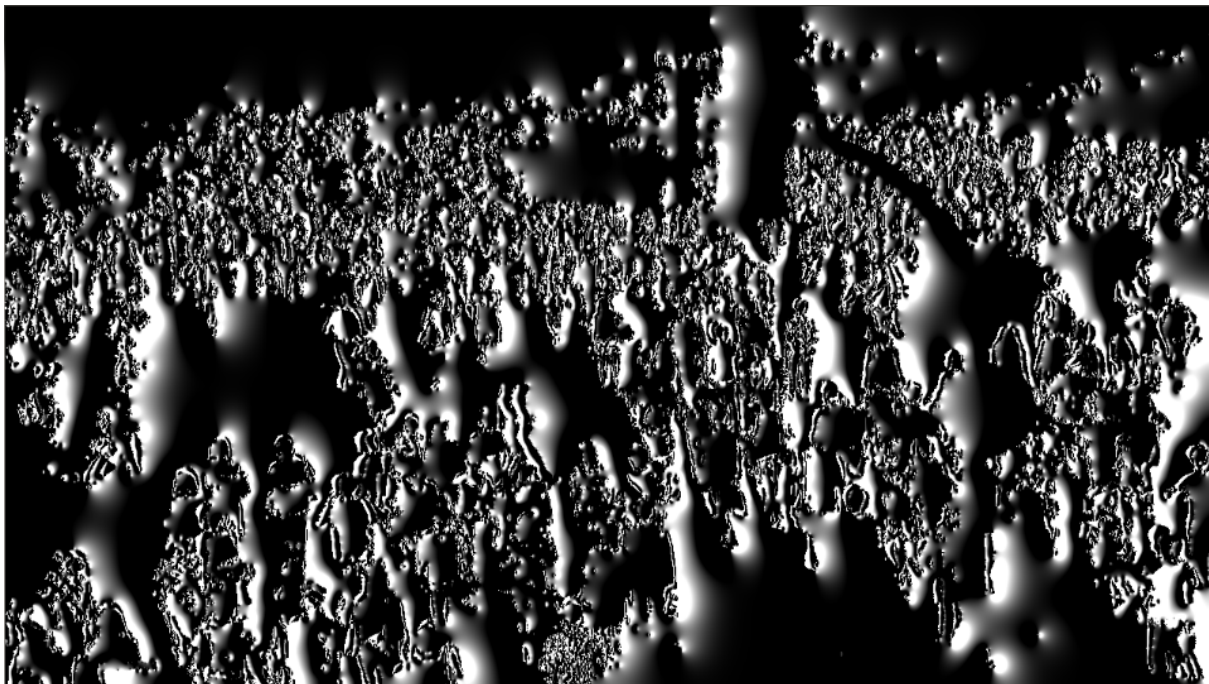


Figura 2.5: Componente horizontal de velocidade da Figura 2.3.

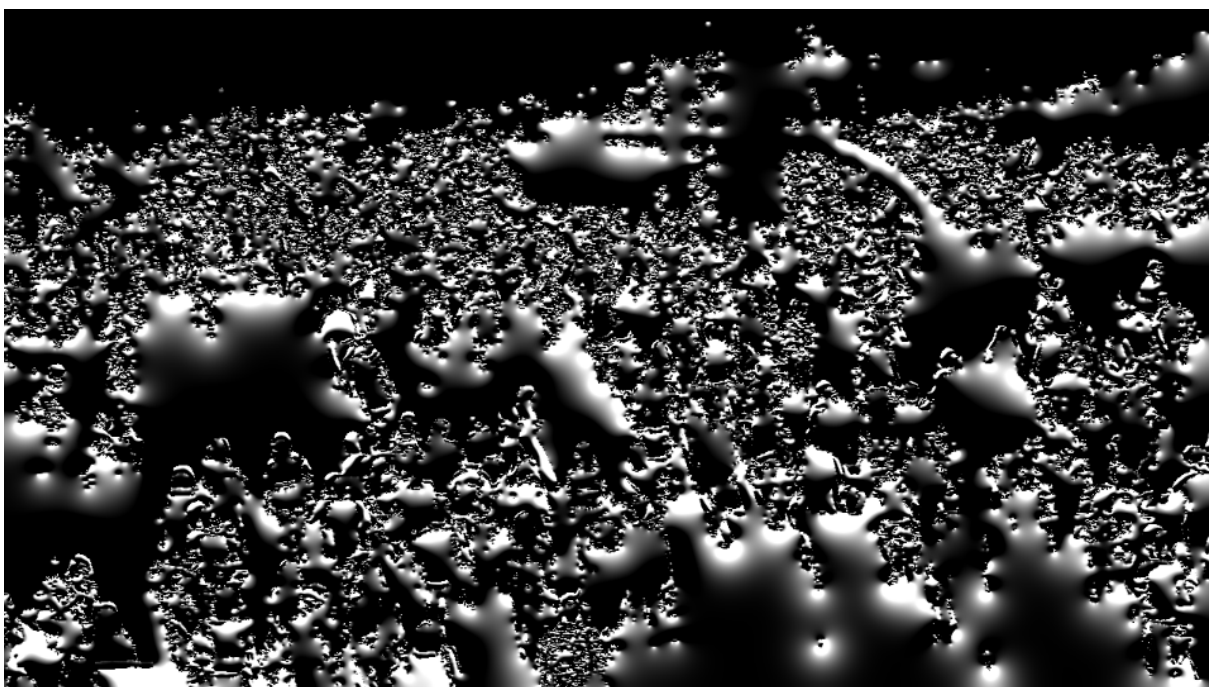


Figura 2.6: Componente vertical de velocidade da Figura 2.3.

Cada uma das derivadas parciais é estimada a partir das médias das quatro primeiras diferenças obtidas pelas medições adjacentes.

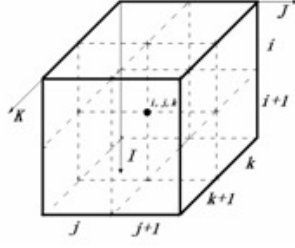


Figura 2.7: Derivadas parciais aproximadas da imagem, retirado de [5] .

$$\begin{aligned}
I_x &\approx \frac{1}{4} \{ I_{i,j+1,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i+1,j,k} + \\
&I_{i,j+1,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i+1,j,k+1} \} \\
I_y &\approx \frac{1}{4} \{ I_{i+1,j,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i,j+1,k} + \\
&I_{i+1,j,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i,j+1,k+1} \} \\
I_t &\approx \frac{1}{4} \{ I_{i,j,k+1} - I_{i,j,k} + I_{i+1,j,k+1} - I_{i+1,j,k} + \\
&I_{i,j+1,k+1} - I_{i,j+1,k} + I_{i+1,j+1,k+1} - I_{i+1,j+1,k} \}
\end{aligned} \tag{2.4}$$

O cálculo da restrição adicional da velocidade é resultado a partir da hipótese que o campo de velocidade é suave. O fator de ponderação  $\alpha^2$  é introduzido para associar a magnitude do erro com a quantização de erros e ruído.

Os valores estimados para as componentes de velocidade de  $u_{k+1}$  e  $v_{k+1}$  são obtidos de

$$u^{k+1} = \bar{u}^k - \frac{I_x [I_x \bar{u}^k + I_y v^k + I_t]}{(\alpha^2 + I_x^2 + I_y^2)} \tag{2.5}$$

$$v^{k+1} = \bar{v}^k - \frac{I_y [I_x \bar{u}^k + I_y v^k + I_t]}{(\alpha^2 + I_x^2 + I_y^2)} \tag{2.6}$$

Na Equação 2.5 e Equação 2.6,  $\bar{u}^k$  e  $\bar{v}^k$  são as velocidades médias aproximadas da Laplaciana dos padrões de brilho na iteração  $k$ , na qual os valores de pixels vizinhos são ponderados com a máscara mostrada em [13].

## 2.4 Transformada de Fourier

A Série de Fourier, idealizada pelo matemático francês Jean Baptiste Joseph Fourier em 1807, é um somatório de senos e cossenos que pode representar qualquer função periódica,

como exemplificado na Figura 2.8.

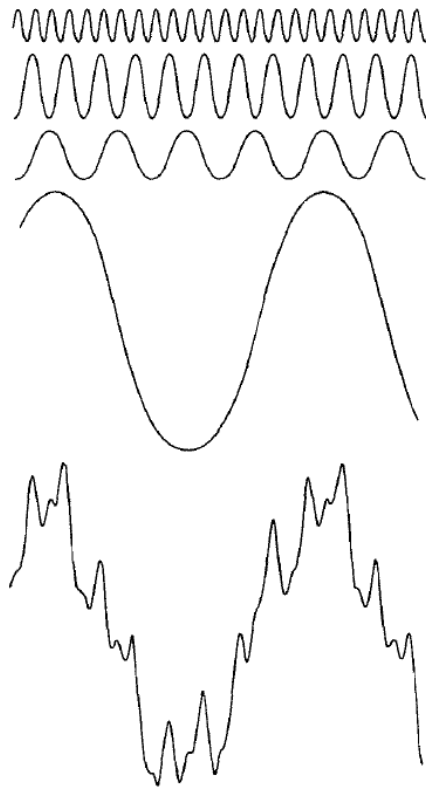


Figura 2.8: Uma função periódica representada pela soma de senos e cossenos, retirada de [11].

A Série de Fourier em termos de variáveis complexas é dada pela Equação 2.7[11]. As magnitudes dos termos complexos da fórmula são os espectros de Fourier.

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{j\frac{2\pi n}{T}t} \quad (2.7)$$

onde

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-j\frac{2\pi n}{T}t} dt \quad (2.8)$$

para  $n = \pm 1, \pm 2, \dots$

Uma característica e vantagem em comum entre a Série e a Transformada de Fourier é a capacidade de retorno à função original a partir da inversa da Série ou da Transformada. Ou seja, não há perda de informações.

### 2.4.1 Impulso

Um impulso unitário de uma função contínua, representado pela variável  $\delta$  é definido por

$$\delta(t) = \begin{cases} \infty & \text{se } t = 0 \\ 0 & \text{se } t \neq 0 \end{cases} \quad (2.9)$$

e deve satisfazer

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (2.10)$$

Para funções discretas, o impulso tem valor máximo de 1 no tempo  $x = 0$ , e a sua equivalência se dá por

$$\sum_{x=-\infty}^{\infty} \delta(x) = 1 \quad (2.11)$$

Ao multiplicar-se a Equação 2.12 por um impulso  $x = x_0$ , tem-se a filtragem por *sifting*, peneiramento, dada pela Equação 2.13.

$$\sum_{x=-\infty}^{\infty} f(x) \delta(x - x_0) = f(x_0) \quad (2.12)$$

Um trem de impulsos é dado pelo somatório de infinitos impulsos. A Transformada de Fourier para esse trem de impulsos, Equação 2.13, é imprescindível, pois é o mesmo que obter a soma das Transformadas dos componentes exponenciais individuais[11]. A equação é derivada da Série de Fourier.

$$\frac{1}{\Delta T} \sum_{n=-\infty}^{\infty} \delta(\mu - \frac{n}{\Delta T}) \quad (2.13)$$

## 2.4.2 Convolução

Para que haja uma convolução, é preciso ter duas funções. Uma delas é escolhida para ser rotacionada em 180° de acordo com sua origem e vai deslizando por cima da outra função intacta. Durante o deslizamento, ocorre uma soma de produtos ponto a ponto, no caso da função ser discreta.

Quando a função é contínua, pode-se afirmar que o produto da sua Transformada de Fourier no domínio da frequência ou a convolução das duas funções no domínio espacial possuem o mesmo resultado[11].

## 2.4.3 Sampling

Para que uma função contínua seja convertida para uma função discreta, ela é multiplicada por um trem de impulsos. Esse processo de *sampling* ocorre na digitalização do sinal.

O teorema de *sampling* afirma que "uma função de banda limitada pode ser completamente recuperada de um conjunto de suas amostras se essas forem adquiridas a uma taxa que excede o dobro da mais alta frequência contida na função."

Quando a taxa é menor do que esse dobro, ocorre o serrilhamento (*aliasing*).

#### 2.4.4 2-D

Até então, todas as fórmulas apresentadas tratam de apenas uma variável. Quando há duas variáveis, as equações passam a ser 2-D.

Para o impulso, tem-se as mesmas propriedades de apenas uma variável, mas com a inserção de uma variável a mais.

Para o *sampling*, a multiplicação é realizada por trens de impulso 2-D. O efeito de serrilhamento permanece, pois nenhuma função é amostrada infinitamente[11].

#### 2.4.5 Transformada Discreta de Fourier em 2-D

A Transformada Discreta de Fourier(DFT) em 2-D é dada pela Equação 2.14[3].

$$F(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) e^{-i2\pi(\frac{ki}{N} + \frac{lj}{N})} \quad (2.14)$$

Algumas propriedades da DFT são: a conversão do domínio espacial para o domínio da frequência; a capacidade de rotacionar o período para a origem, com o deslocamento da DFT.

Por ser geralmente complexa, essa pode ser facilmente expressa em magnitude e fase. Para o processamento de imagens, a parte mais importante é a magnitude, pois ela contém todas as informações necessárias sobre a estrutura geométrica das imagens[3].

As Figuras 3.5 a 3.6 dão um exemplo visual da Transformada Discreta de Fourier para uma imagem 2-D.

#### 2.4.6 Aplicações

Um filtro no domínio da frequência é dado pela multiplicação da convolução com a Transformada de Fourier da imagem de entrada. Dada uma imagem de entrada denominada  $f(x, y)$ , a equação de filtragem é

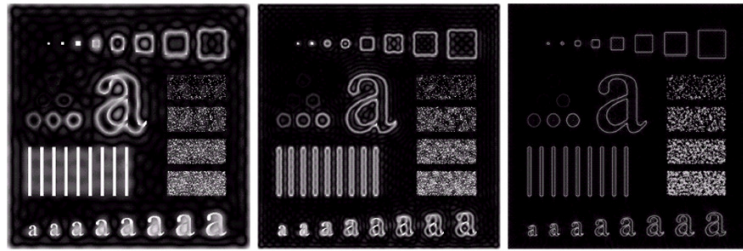
$$g(x, y) = \mathfrak{F}^{-1}[H(u, v)F(u, v)] \quad (2.15)$$

Onde  $\mathfrak{S}^{-1}$  é a Transformada Discreta de Fourier Inversa,  $H(u, v)$  é uma função de filtro,  $F(u, v)$  é a Transformada Discreta de Fourier da imagem de entrada e  $g(x, y)$  é a imagem filtrada.

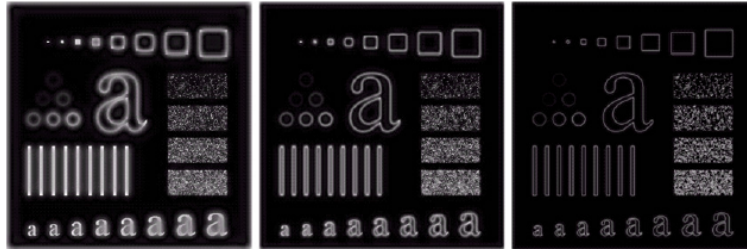
Os filtros mais comuns são passa-alta, onde cortam as baixas frequências e mantem as altas frequências; passa-baixa, onde cortam as altas frequências e mantem as baixas frequências ou passa-faixa, que permitem que uma faixa definida de frequência passe.

Os filtros passa-baixa suavizam a imagem, pois as altas frequências compõem os detalhes da imagem, podendo esta perder a nitidez.

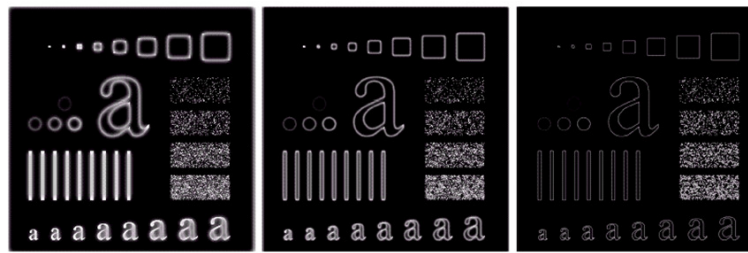
Já os filtros passa-alta definem melhor os detalhes da imagem, fazendo com que estes sejam evidenciados, como apresentado na Figura 2.9.



(a) Filtro ideal



(b) Filtro Butterworth



(c) Filtro Gaussiano

Figura 2.9: Exemplos de filtros passa-alta

A Figura 2.9 apresenta uma comparação entre os filtros passa-altas.

# Capítulo 3

## Metodologia Proposta

A Figura 3.1 apresenta o fluxograma do método utilizado para a comprovação da hipótese e para o processo de detecção de eventos anormais em câmeras de videomonitoramento. Este fluxograma é a modelagem do descritor de eventos anormais em multidões proposto e com seu resultado, obtem-se as informações necessárias para a criação de tal descritor.

Devido ao fluxo ótico, é necessário que cada *frame* passe por todas as etapas apresentadas no fluxograma a partir do segundo *frame* de vídeo. Aplicou-se o fluxo ótico Horn e Schunck [13]. A partir da estimativa de densidade do fluxo ótico, os componentes horizontais e verticais de velocidade são avaliados.

A transformada discreta de Fourier em duas dimensões é aplicada em ambos os campos de movimento estimado, horizontal e vertical. A transformada discreta de Fourier divide o campo de movimento em componentes de alta e baixa frequência em regiões específicas. Considerando essa informação, dois métodos espaciais de corte distintos são propostos, um retangular e um quadricular, para extrair as informações dos componentes de alta e baixa frequência do campo de movimento estimado.

Para este trabalho, as imagens usadas são vídeos de câmeras sem calibração, sendo as análises baseadas em vídeos filmados por câmeras de segurança e nenhum conhecimento prévio do evento.

### 3.1 Sequência de Imagens de Entrada

Para esse projeto, apenas vídeos de câmeras de segurança foram utilizados. Um vídeo de segurança é, normalmente, capturado de uma câmera fixa que pode ser encontrada em diversos pontos da cidade nos dias atuais [16]. Nenhuma outra informação da cena é usada e toda a informação de processamento de imagem é feita no *frame* atual do vídeo, definido no índice temporal  $t + 1$  e comparado com o *frame* anterior no tempo  $t$ .

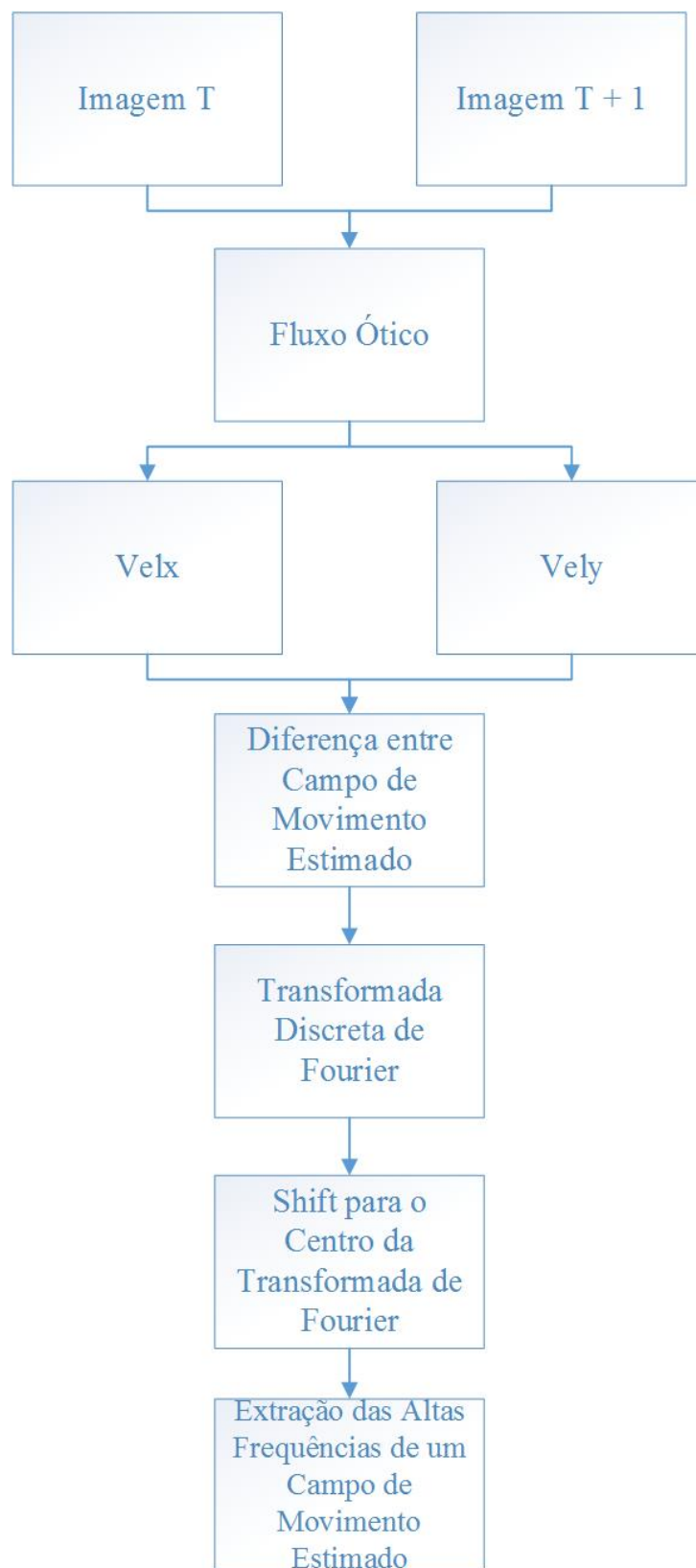


Figura 3.1: Fluxograma metodológico.



## 3.2 Fluxo Ótico

O fluxo ótico aplicado no processo é a técnica diferencial de Horn e Schunck [13], explicado na Seção 2.3.1, Capítulo 2.

## 3.3 Diferença entre Campo de Movimento Estimado

Após a avaliação do fluxo ótico, dois componentes de velocidade são estimados nas direções horizontais ( $Vel_x$ ) e verticais ( $Vel_y$ ), Figuras 3.3 a 3.4. Esses componentes possuem o mesmo tamanho da imagem de entrada e estes descrevem o campo de movimento no domínio da imagem.

Para o processo de detecção de eventos anormais em multidões, a diferença absoluta do fluxo estimado de  $k + 1$  e  $k$  foi calculada, pois em muitos casos as mudanças de alta luminância nas câmeras de segurança afetam diretamente a técnica proposta de fluxo ótico.



Figura 3.2: Vídeo original.

## 3.4 Transformada Discreta de Fourier

Para a extração das componentes de alta frequência, a Transformada Discreta de Fourier é utilizada. De acordo com a Equação 2.14, os espectros de alta e baixa frequência

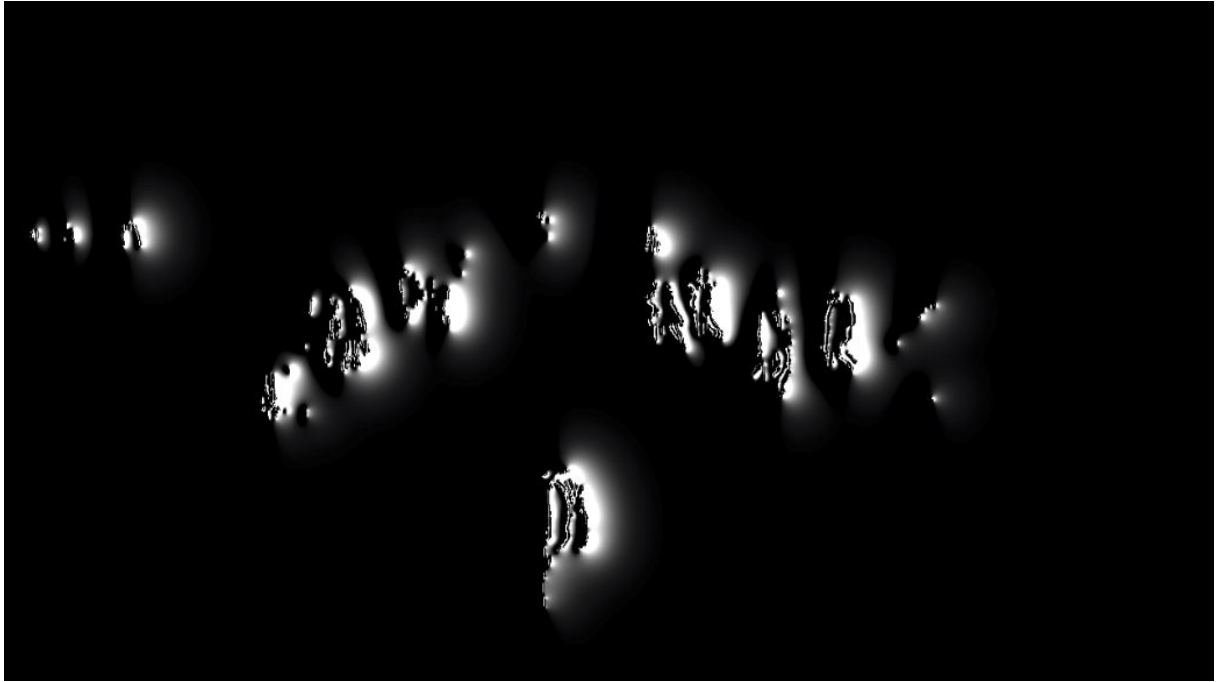


Figura 3.3: Componente horizontal do campo de movimento da Figura 3.2.

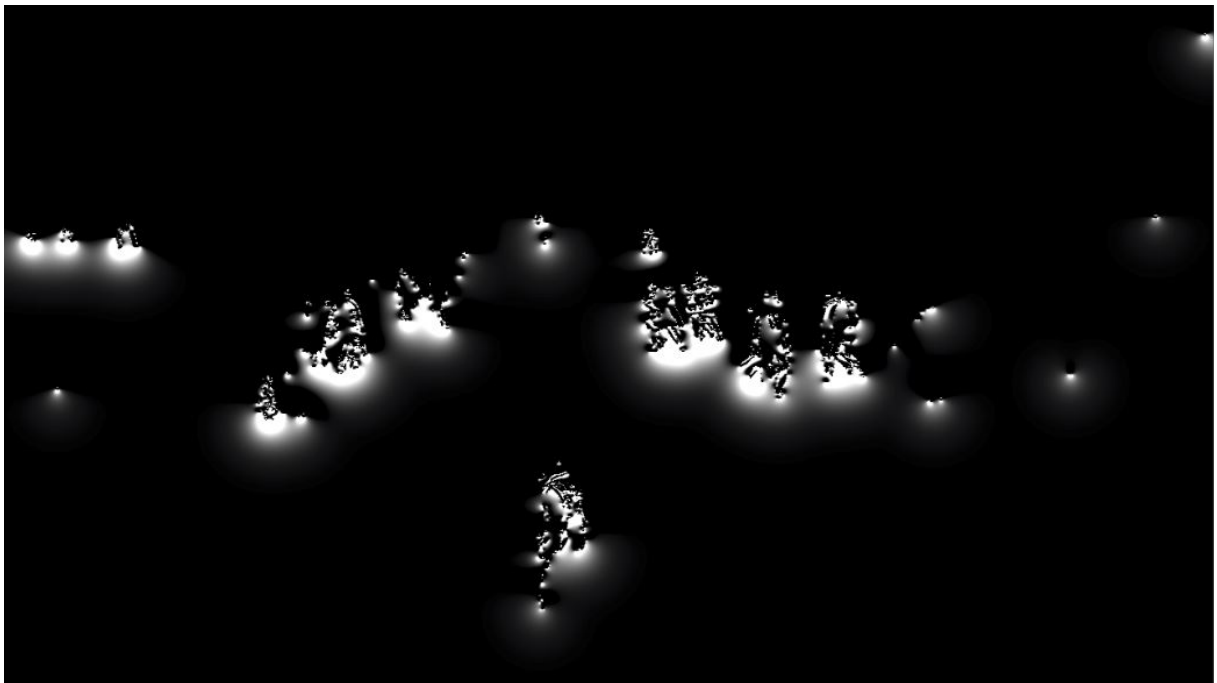


Figura 3.4: Componente vertical do campo de movimento da Figura 3.2.

são calculados depois da transformação e adquiridos de cada componente sinusoidal e invariante da informação espacial da imagem.

Um exemplo da transformada discreta dos componentes horizontais e verticais está nas Figuras 3.5 a 3.6. Para fins de visualização, as magnitudes foram normalizadas, mas para a resolução efetiva do problema proposto não há normalização.

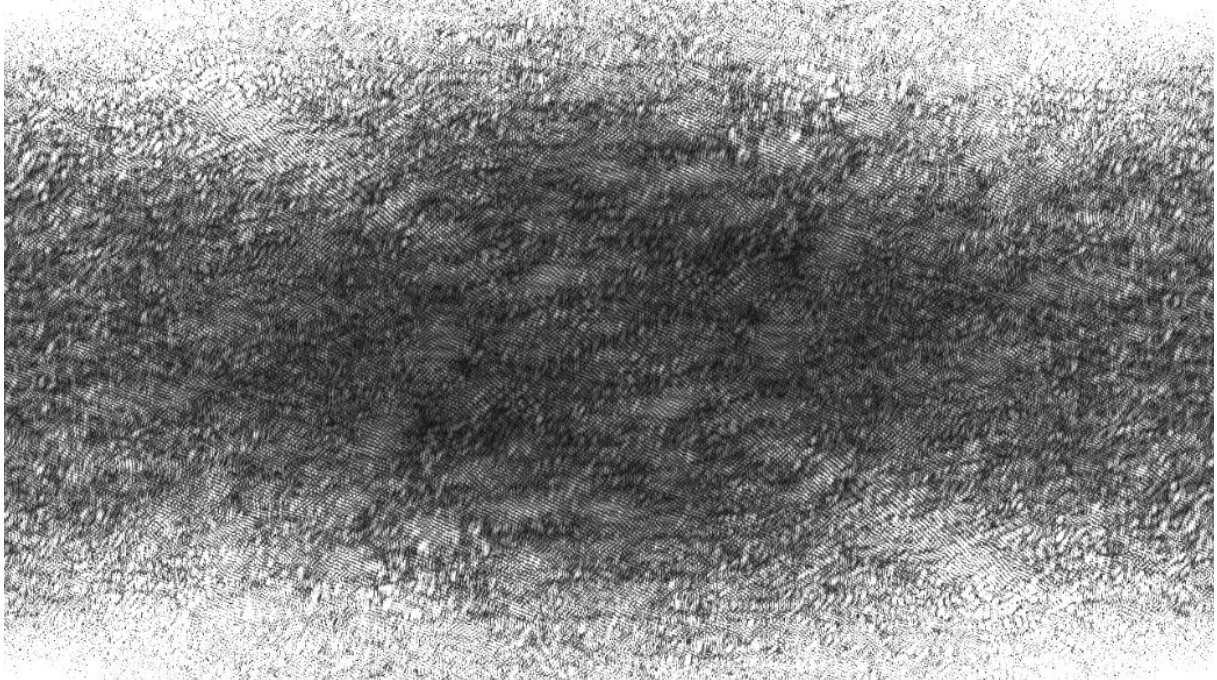


Figura 3.5: Magnitudes da componente horizontal do fluxo ótico da Figura 3.2.

### 3.5 Deslocamento da Transformada de Fourier

Para um melhor resultado no estágio de classificação, depois da transformada discreta de Fourier todos os componentes de campo de movimento estimados foram deslocados para o centro da imagem, Figuras 3.7 a 3.8.

Quando todas as magnitudes são deslocadas para o centro, é possível realocar espacialmente as posições das altas e baixas frequências. Dessa maneira, as altas frequências estão nas bordas do espectro de magnitude e as baixa frequências estão no meio do espectro de magnitude.

O deslocamento é essencial para que a classificação baseada em características espaciais dos componentes de alta frequência possa ser implementado.



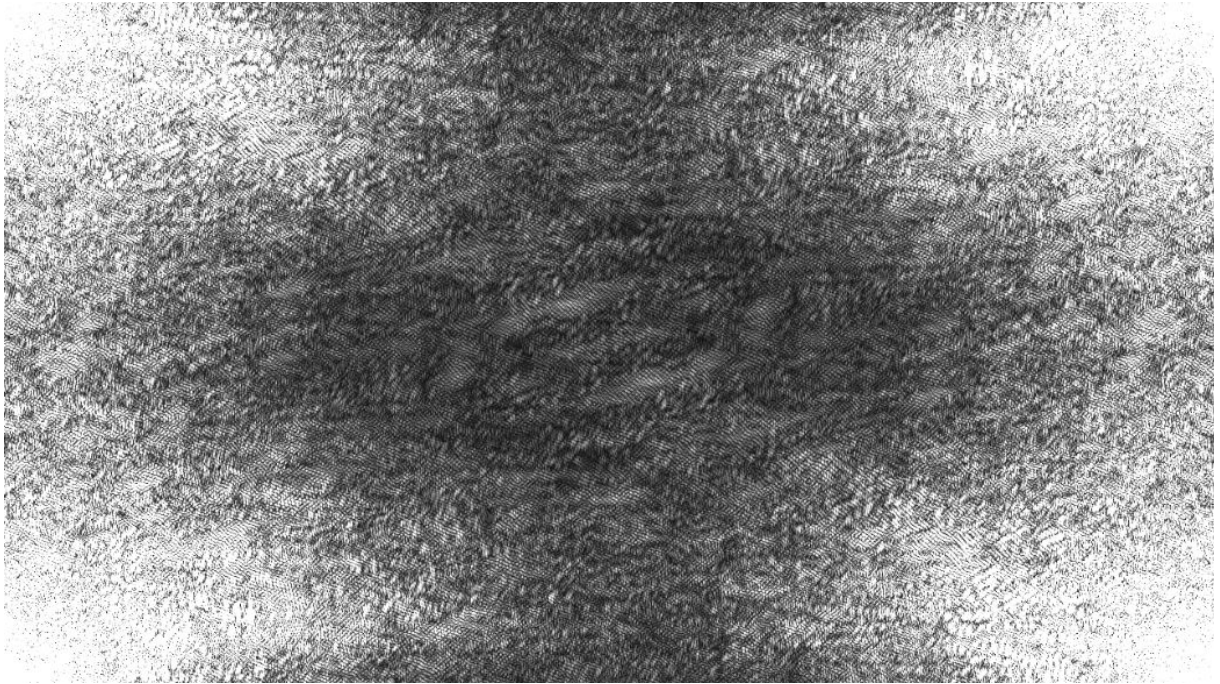


Figura 3.6: Magnitudes da componente vertical do fluxo ótico da Figura 3.2.



Figura 3.7: Deslocamento para o centro da componente horizontal do fluxo ótico da Figura 3.2.



Figura 3.8: Deslocamento para o centro da componente vertical do fluxo ótico da Figura 3.2.

### 3.6 Extração das Componentes de Alta Frequência do Campo de Movimento Estimado

Para fazer a detecção de eventos anormais em multidões em câmeras de segurança assume-se a hipótese de que quando há uma briga, por exemplo, em um ambiente com muitas pessoas grande parte das componentes de altas frequências tem uma mudança de amplitude, ou há como constatar essas variações, na dispersão espacial quando comparada com a dispersão original desses componentes.

Para realizar a construção do descritor, duas abordagens de características espaciais (retângulos e aros retangulares) baseadas apenas na análise da dispersão espacial foram elaboradas, calculando as mudanças em regiões onde as altas frequências estão localizadas e descartando as regiões das baixas frequências que estão localizadas no espectro.

Os cálculos das regiões se dão pelas médias da soma de todos os elementos da imagem dentro daquela região específica. Esta abordagem é simples e em caso de sucesso será suficiente para indicar a comprovação da hipótese.

Para ambos os métodos de extração, a variável número de cortes é dinâmica e aceita qualquer número inteiro desejado.

### 3.6.1 Extração por Retângulos

A extração por retângulos foi idealizada de acordo com a Figura 3.9. A soma dos elementos é feita em cada divisão e a técnica é executada em ambos os componentes, ou seja, nas magnitudes deslocadas de  $Vel_x$  e  $Vel_y$ .

Na Figura 3.9,  $L$  é a largura e  $NdC$  denota o número de cortes. A Figura 3.10 é uma representação da intensidade da soma dos elementos, assim feita para fins de apresentação visual do resultado esperado.

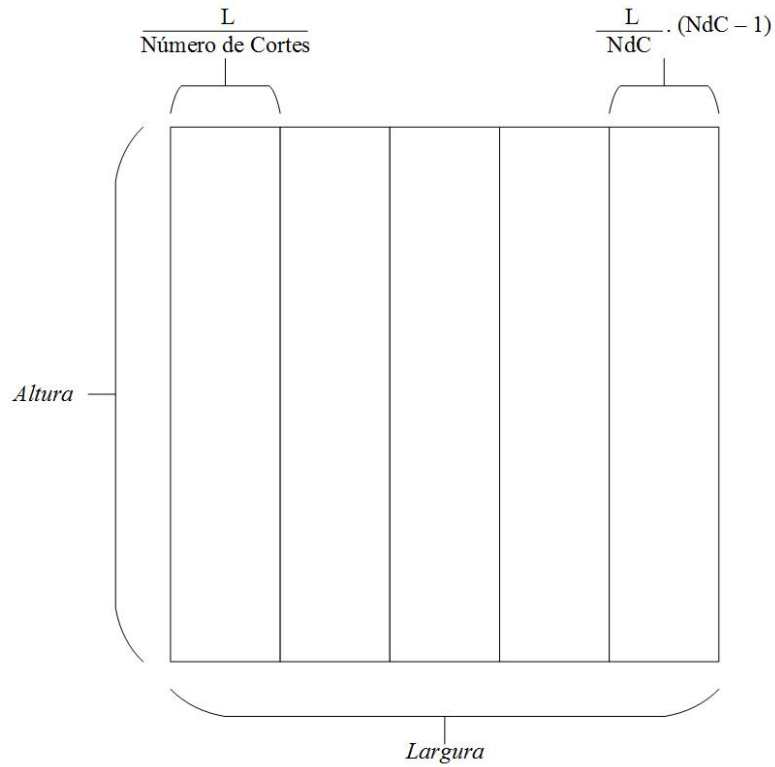


Figura 3.9: Análise por retângulos.

### 3.6.2 Extração por Arcos Retangulares

Pela disposição das frequências após o deslocamento da Transformada de Fourier, a extração de altas frequências em formatos de arcos retangulares é a consideração mais natural a se fazer. De acordo com a Figura 3.11, cada divisão é considerada de forma distinta, ou seja, para cada arco retangular selecionado os outros arcos retangulares são descartados, colocando-se a cor preta nesses locais indesejados e a soma dos elementos da imagem naquele pedaço de corte é desempenhada.



Figura 3.10: Análise do componente vertical com cinco retângulos da Figura 3.2.

Para a Figura 3.11,  $A$  simboliza a altura,  $NdC$  é o número de cortes e  $L$  representa a largura. A Figura 3.12 apresenta uma aplicação real realçada, para melhorar a visualização, desse método para três cortes.

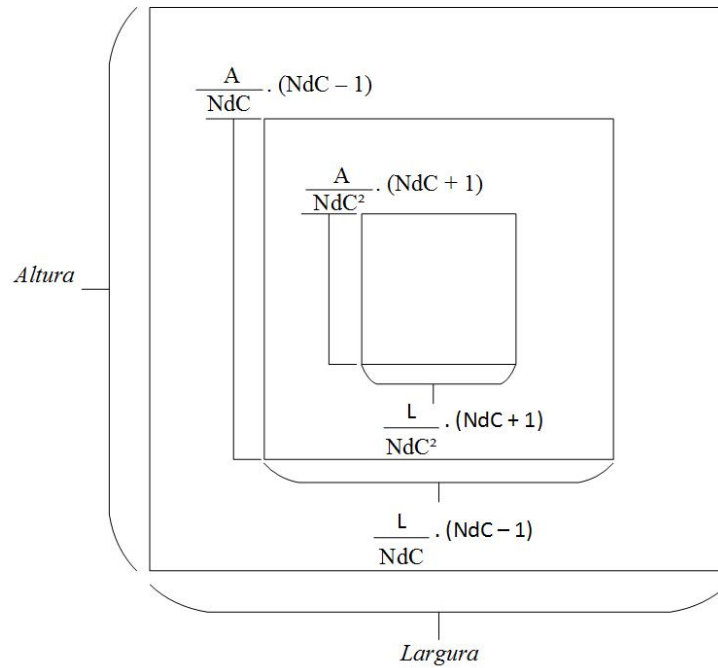


Figura 3.11: Análise por arcos retangulares.



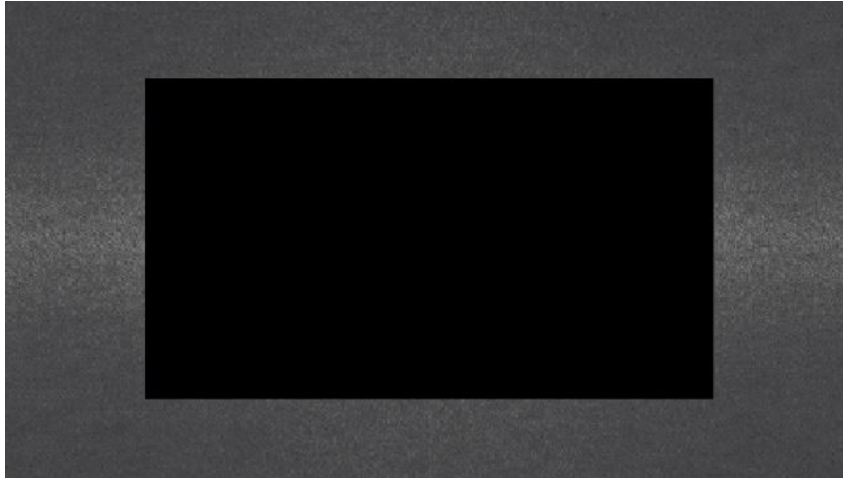


Figura 3.12: Análise com três arcos retangulares da Figura 3.2.

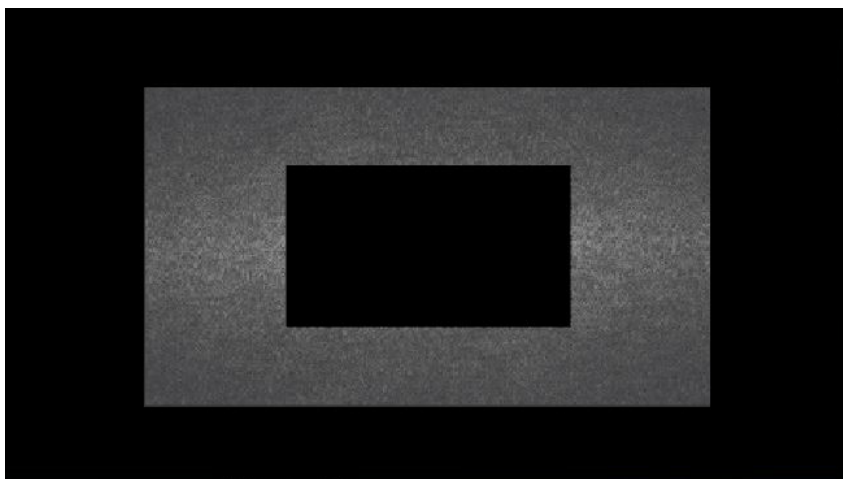


Figura 3.13: Análise com três arcos retangulares da Figura 3.2.



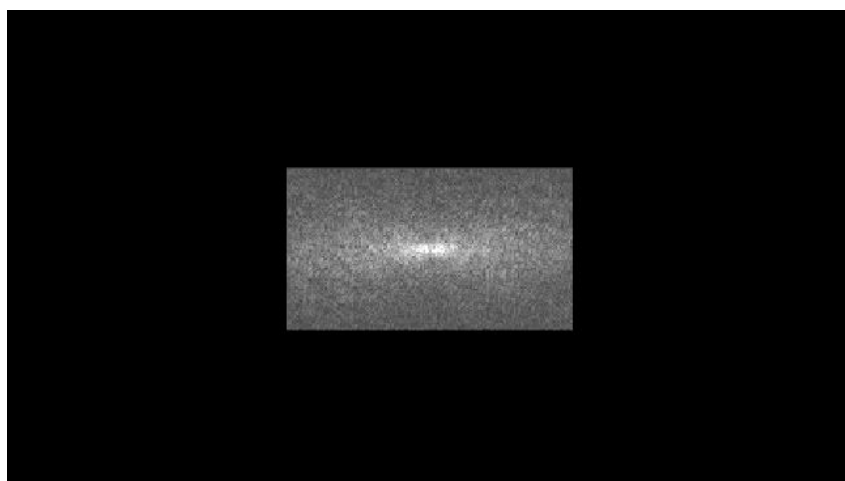


Figura 3.14: Análise com três arcos retangulares da Figura 3.2.

# Capítulo 4

## Resultados

Para fins de validação da técnica construída para confirmar a hipótese, uma base de dados composta por vinte vídeos foi montada. Todos esses vídeos serão apresentados nas Seções 4.1.1 e 4.1.2.

O código foi desenvolvido usando a biblioteca para processamento de imagens OpenCV e para os resultados gráficos, a ferramenta utilizada foi o MATLAB.

A máquina usada para a execução do código possui as seguintes especificações:

- Memória: 8GB;
- Processador: Intel Core i7-4702HQ;
- GPU: NVIDIA GeForce GTX 870M;
- Sistema Operacional: Linux Ubuntu 12.04 LTS 64-bit.

### 4.1 Base de Dados

Vinte vídeos foram selecionados para compor a base de dados desse projeto. Todos eles foram retirados da Internet e configuram uma mesma característica: todos são de câmeras de segurança, logo, todos utilizam uma câmera fixa. Além disso, a duração deles é, em média, de trinta segundos.

Desses vinte vídeos, dez são de situações de multidões em eventos normais, por exemplo: pessoas andando nas ruas e carros passando. Os outros dez vídeos são de agitação em multidões, cenas de violência, em sua maioria. Configuram violência casos como: brigas em show, brigas em bares, entre outros.

Portanto, divide-se essa base de dados em duas classes distintas: Vídeos calmos e vídeos agitados, como descrito a seguir.

Os nomes dos vídeos, suas resoluções, durações e FPS estão descritos na Tabela 4.1 e na Tabela 4.2.

### 4.1.1 Vídeos Calmos

Os vídeos calmos utilizados estão nas Figuras 4.1 a 4.5.



Figura 4.1: À esquerda, pessoas caminhando pela Abbey Road. À direita, turistas em Wall Street.



Figura 4.2: À esquerda, pedestres em Dublin. À direita, imagem da Times Square.



Figura 4.3: À esquerda, transeuntes na Times Square. À direita, rua na Itália.



Figura 4.4: À esquerda, movimentação normal na Times Square. À direita, pedestres na Quinta Avenida.



Figura 4.5: À esquerda, praia do Rio de Janeiro. À direita, pessoas atravessando um cruzamento em Shibuya.

Tabela 4.1: Descrição Completa dos Vídeos Calmos Utilizados.

Nome do Vídeo	Amostra	Resolução	Duração	FPS
abbey	À esquerda da Figura 4.1	1280x720	31s	30
wallst	À direita da Figura 4.1	1280x720	30s	30
dublin	À esquerda da Figura 4.2	480x272	34s	25
times	À direita da Figura 4.2	480x272	30s	25
times2	À esquerda da Figura 4.3	480x272	30s	25
italy	À direita da Figura 4.3	1280x720	30s	30
times3	À esquerda da Figura 4.4	854x480	30s	25
fifth	À direita da Figura 4.4	480x272	31s	25
rio	À esquerda da Figura 4.5	480x272	31s	25
shibuya	À direita Figura 4.5	1280x720	30s	25



### 4.1.2 Vídeos Agitados

Os vídeos da classe agitados estão nas Figuras 4.6 a 4.10



Figura 4.6: À esquerda, tumulto em show. À direita, ocorrência de uma briga em um show.



Figura 4.7: À esquerda, briga em um supermercado. À direita, ocorrência de uma briga em um show.



Figura 4.8: À esquerda, briga entre manifestantes. À direita, briga em um saguão.



Figura 4.9: À esquerda, pedestres em fuga. À direita, ocorrência de uma briga em um bar.



Figura 4.10: À esquerda, tumulto em um cassino. À direita, briga em uma loja de conveniência.

Tabela 4.2: Descrição Completa dos Vídeos Agitados Utilizados.

Nome do Vídeo	Amostra	Resolução	Duração	FPS
crowdwave	À esquerda da Figura 4.6	480x272	29s	25
show	À direita da Figura 4.6	480x272	30s	25
fight4	À esquerda da Figura 4.7	352x240	34s	25
axe	À direita da Figura 4.7	1280x720	32s	29
fight	À esquerda da Figura 4.8	854x480	30s	25
fight5	À direita da Figura 4.8	640x360	30s	25
kiev	À esquerda da Figura 4.9	1920x1080	28s	50
pub	À direita da Figura 4.9	854x480	30s	60
fight7	À esquerda da Figura 4.10	480x272	31s	25
police	À direita da Figura 4.10	854x480	31s	25

## 4.2 Testes Realizados

Os vídeos foram submetidos a dois tipos diferentes de testes. O primeiro teste foi processar todos os vídeos de todas as classes para o método de extração retangular apresentado na Seção 3.6.1. Todos executaram com variados números de cortes: 3, 5, 7 e 11. O segundo teste executou o código para a extração por arcos retangulares, Seção 3.6.2. Os números de corte também foram 3, 5, 7 e 11.

Quanto maior o número de cortes, maior será a precisão das médias das somas, pois uma faixa mais estreita de frequência será calculada. Os números escolhidos para a comprovação da hipótese foram propositalmente ímpares, para que a faixa central de baixa frequência pudesse ser totalmente removida.

Para montar os gráficos, uma média simples das faixas de cortes correspondentes às altas frequências estimadas pelo deslocamento – as faixas que se encontravam nas extremidades – de todos os vídeos daquela classe para o número de cortes indicado foi feita.

As Tabelas 4.3 a 4.6, correspondem às coordenadas  $(x, y)$  dos pontos que são graficamente representados. Os vídeos calmos com método de extração retangular estão entre na Tabela 4.3, os vídeos agitados com método de extração retangular estão na Tabela 4.4. Os vídeos calmos com método de extração por arcos retangulares estão na Tabela 4.5 e os agitados na Tabela 4.6.

Os gráficos dos pontos apresentados nas Tabelas 4.3 a 4.6 foram gerados com a ferramenta *MATLAB*. Os pontos vermelhos sempre representam os vídeos agitados, enquanto os pontos azuis descrevem os vídeos calmos.

Os gráficos dos vídeos calmos versus vídeos agitados, fazem uma comparação dos resultados obtidos nos dois gráficos anteriores em um mesmo plano, enquanto os gráficos de energia analisam a energia da alta frequência de cada classe de forma comparativa.

Os gráficos da extração por retângulos estão entre as Figuras 4.11 a 4.14, enquanto os gráficos da extração por arcos retangulares estão entre as Figuras 4.15 a 4.18.

## 4.3 Análise dos Resultados

A *priori*, a técnica de extração por arcos retangulares retornaria melhores resultados por ser a análise mais lógica a se fazer, devido à forma que o deslocamento das magnitudes das componentes de alta frequência do fluxo ótico toma e as definições de frequência que esse deslocamento traz.

As Figuras 4.11 a 4.15 mostram que essa suposição não é de todo verdadeira, visto que com um método mais simples e sem nenhuma melhoria, que é o de extração retangular, também é possível criar uma classificação.

As médias das altas frequências entre a técnica de extração por retângulos e por arcos retangulares variaram em cerca de 10%, fazendo com que os gráficos gerados obtivessem uma certa semelhança.

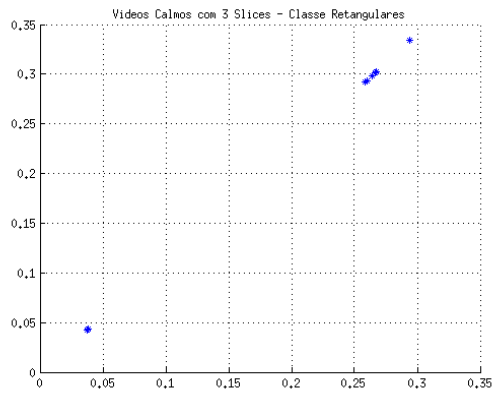
Quanto a mudança de número de cortes, percebe-se que quanto maior o número de corte, mais os pontos convergem em um ponto  $(x, y)$  no espaço de energia, formando uma região definida. Pois quanto mais estreito o corte fica, mais detalhes da alta frequência real são adquiridos, permitindo uma classificação.

Quando colocados em um mesmo plano, como ocorre nas Figuras 4.11 a 4.18, nota-se que há uma nuvem de pontos característicos para cada classe, ou seja, há uma clara diferenciação entre os vídeos agitados e os vídeos calmos. Entretanto, essa distinção não ocorre totalmente em (d), nos gráficos de energia.

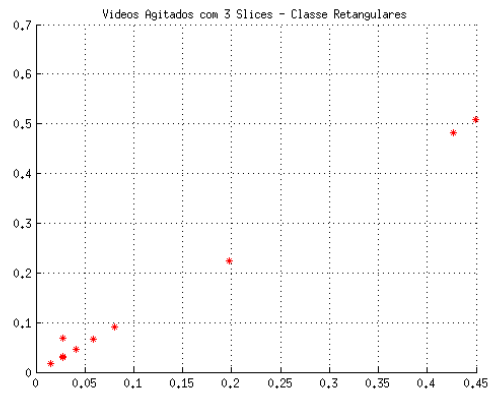
As energias estão correlacionadas para algumas situações em algumas cenas e possuem pouca variação de informação espacial do tipo de extração utilizado.

Ainda assim, a grande maioria dos vídeos agitados localizam-se abaixo da linha dos 0.3. Para fazer a classificação dos resultados, um classificador robusto como uma rede neural pode ser implementado, pois uma classificação simples, como o *threshold*, não conseguirá delimitar se um vídeo possui agitação ou não.

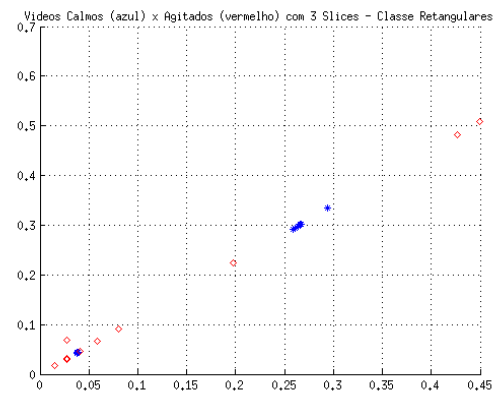




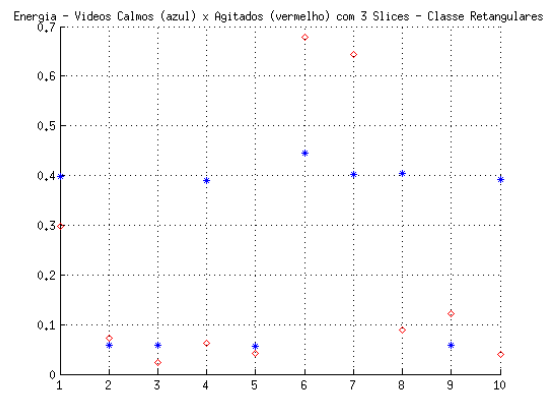
(a) Gráfico dos vídeos calmos para três cortes



(b) Gráfico dos vídeos agitados para três cortes

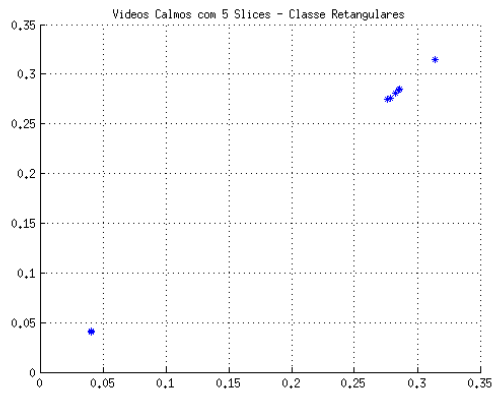


(c) Gráfico dos vídeos calmos de três cortes X  
Gráfico dos vídeos agitados de três cortes

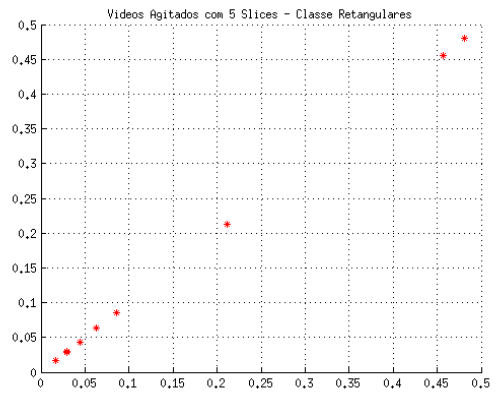


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para três cortes

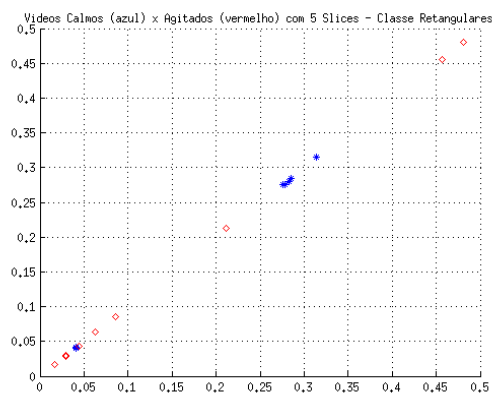
Figura 4.11: Gráficos para três cortes retangulares



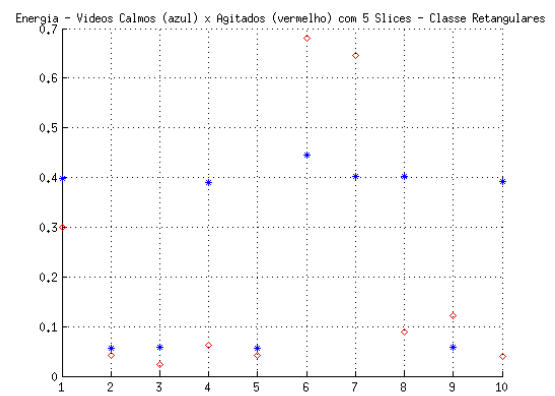
(a) Gráfico dos vídeos calmos para cinco cortes



(b) Gráfico dos vídeos agitados para cinco cortes

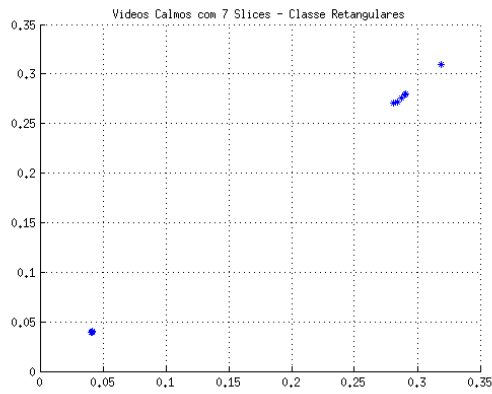


(c) Gráfico dos vídeos calmos de cinco cortes X  
Gráfico dos vídeos agitados de cinco cortes

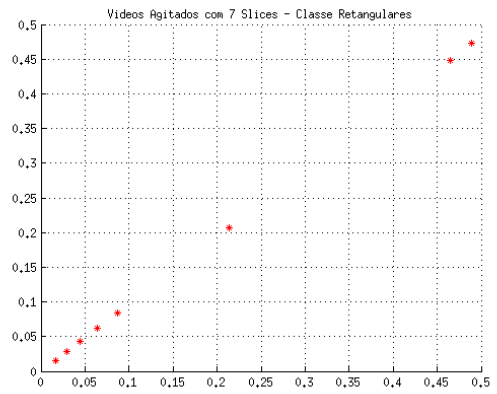


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para cinco cortes

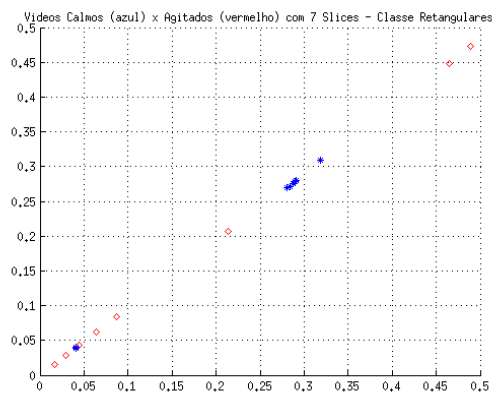
Figura 4.12: Gráficos para cinco cortes retangulares



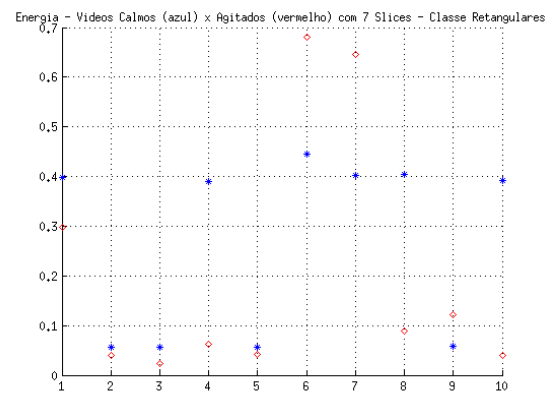
(a) Gráfico dos vídeos calmos para sete cortes



(b) Gráfico dos vídeos agitados para sete cortes

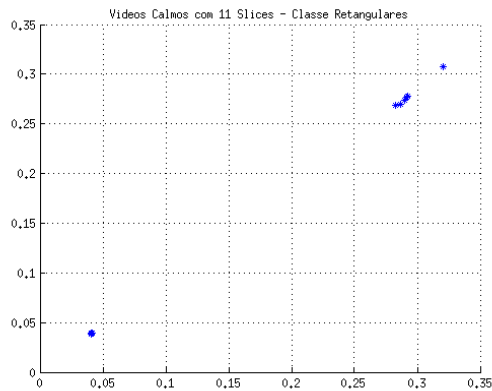


(c) Gráfico dos vídeos calmos de sete cortes X  
Gráfico dos vídeos agitados de sete cortes

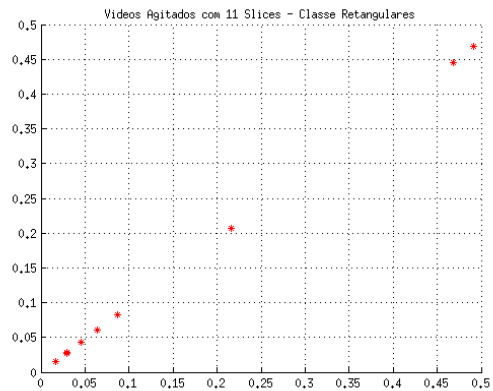


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para sete cortes

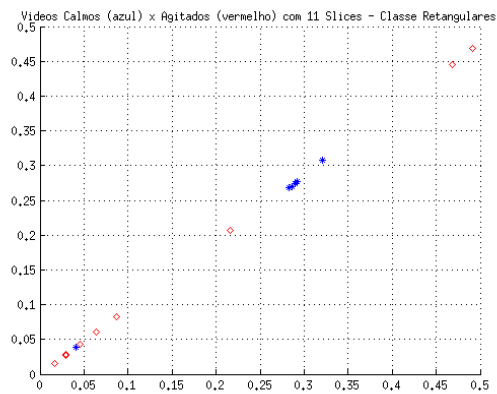
Figura 4.13: Gráficos para sete cortes retangulares



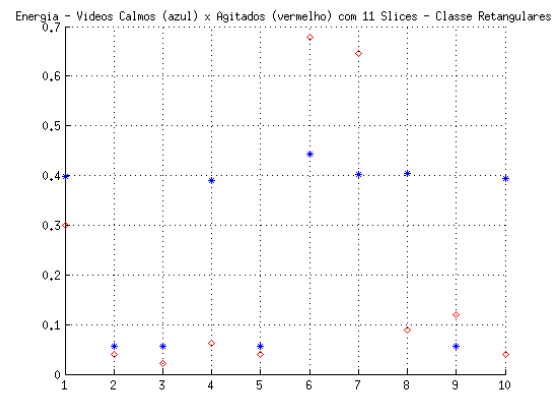
(a) Gráfico dos vídeos calmos para onze cortes



(b) Gráfico dos vídeos agitados para onze cortes

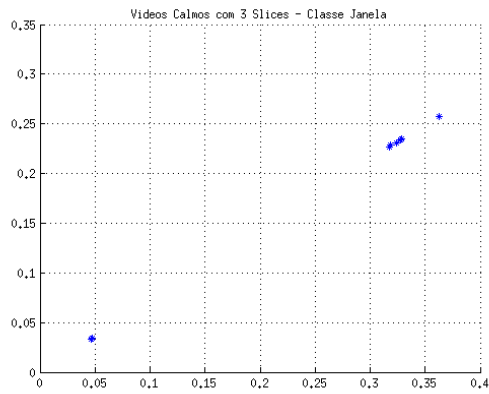


(c) Gráfico dos vídeos calmos de onze cortes X  
Gráfico dos vídeos agitados de onze cortes

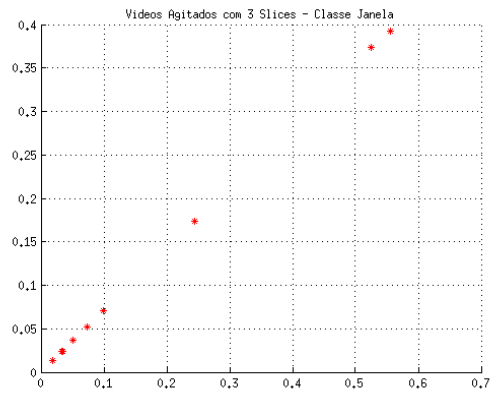


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para onze cortes

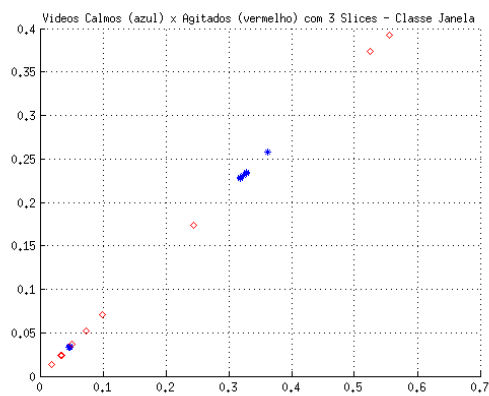
Figura 4.14: Gráficos para onze cortes retangulares



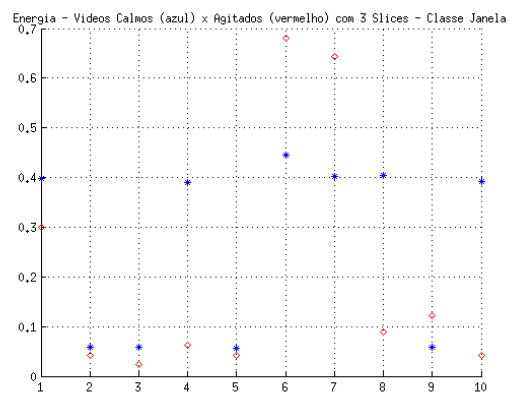
(a) Gráfico dos vídeos calmos para três cortes



(b) Gráfico dos vídeos agitados para três cortes

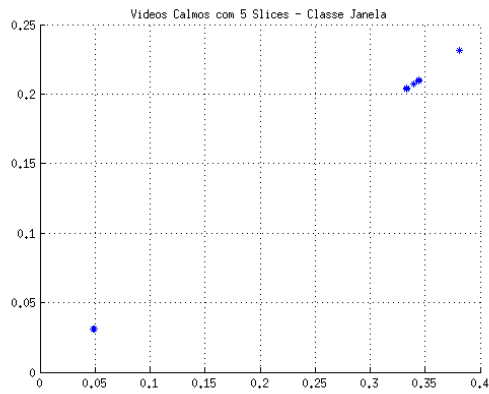


(c) Gráfico dos vídeos calmos de três cortes X  
Gráfico dos vídeos agitados de três cortes

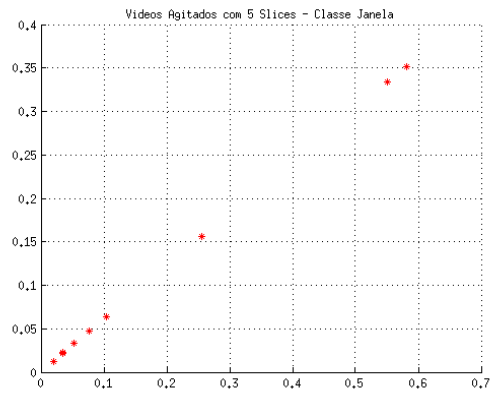


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para três cortes

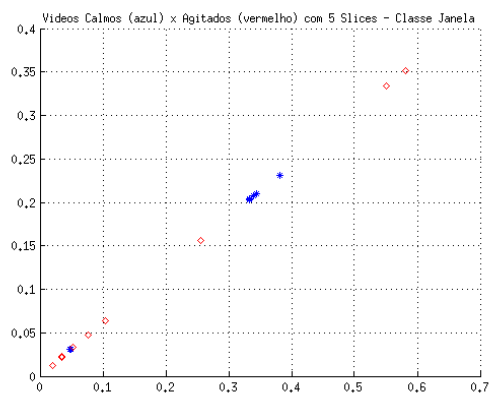
Figura 4.15: Gráficos para três cortes por arcos retangulares



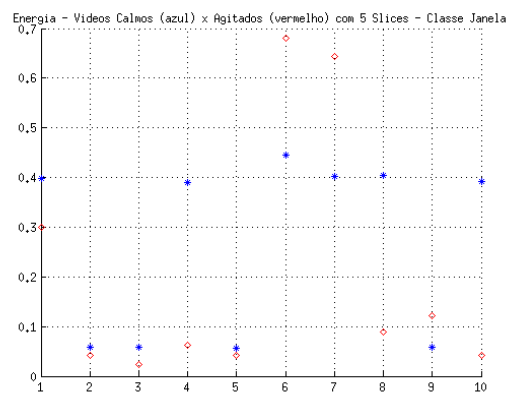
(a) Gráfico dos vídeos calmos para cinco cortes



(b) Gráfico dos vídeos agitados para cinco cortes

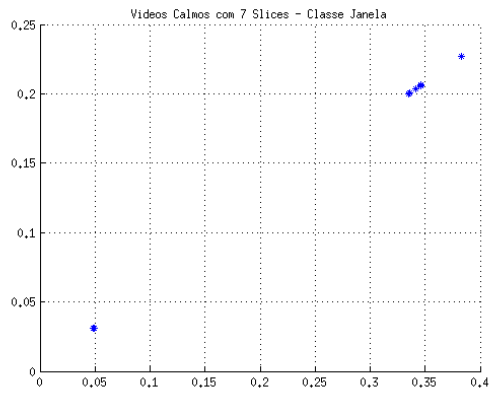


(c) Gráfico dos vídeos calmos de cinco cortes X  
Gráfico dos vídeos agitados de cinco cortes

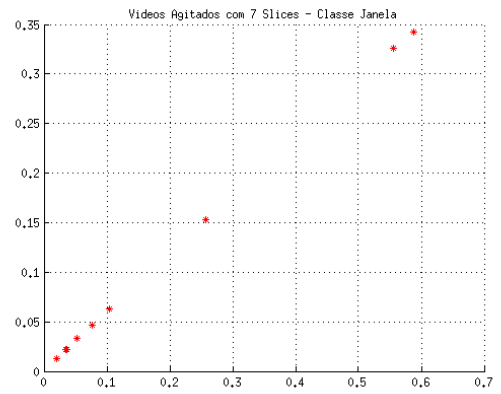


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para cinco cortes

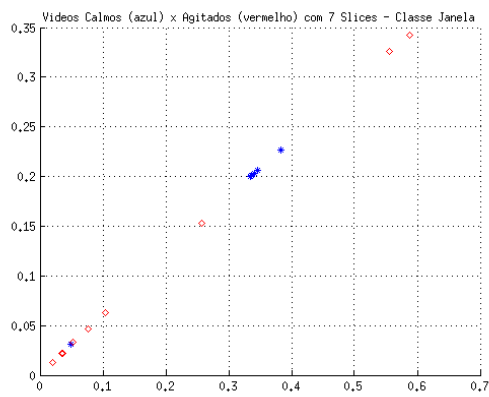
Figura 4.16: Gráficos para cinco cortes por arcos retangulares



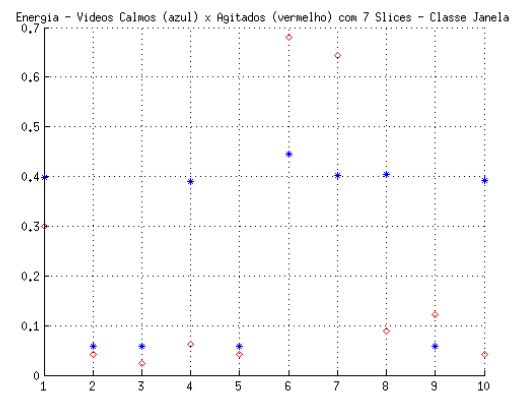
(a) Gráfico dos vídeos calmos para sete cortes



(b) Gráfico dos vídeos agitados para sete cortes

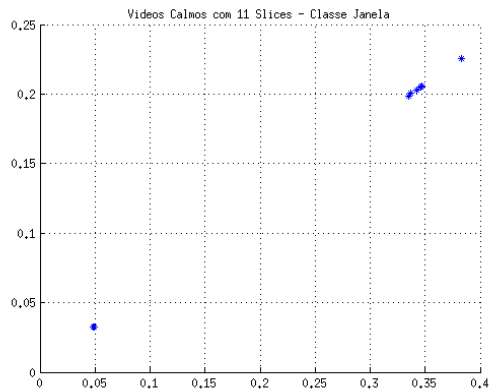


(c) Gráfico dos vídeos calmos de sete cortes X  
Gráfico dos vídeos agitados de sete cortes

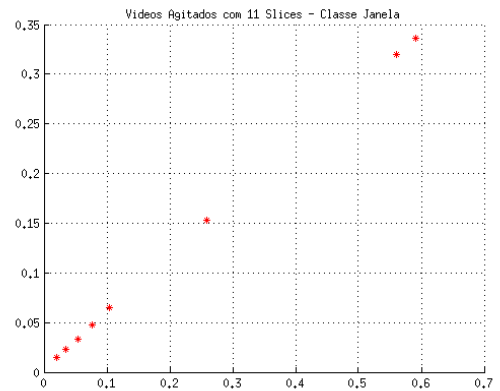


(d) Gráfico de energia dos vídeos calmos X vídeos  
agitados para sete cortes

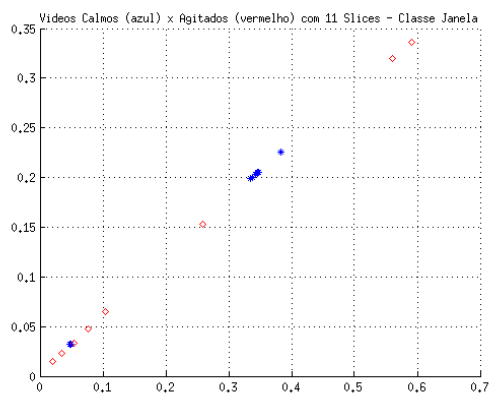
Figura 4.17: Gráficos para sete cortes por arcos retangulares



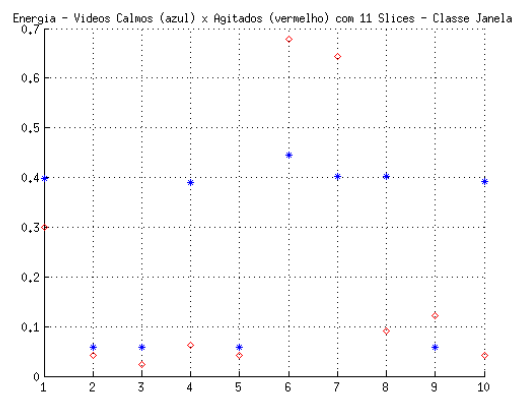
(a) Gráfico dos vídeos calmos para onze cortes



(b) Gráfico dos vídeos agitados para onze cortes



(c) Gráfico dos vídeos calmos de onze cortes X Gráfico dos vídeos agitados de onze cortes



(d) Gráfico de energia dos vídeos calmos X vídeos agitados para onze cortes

Figura 4.18: Gráficos para onze cortes por arcos retangulares



Tabela 4.3: Coordenadas  $(x, y)$  para os vídeos calmos com extração retangular.

Nome do Vídeo	Três Cortes		Cinco Cortes		Sete Cortes		Onze Cortes	
	X	Y	X	Y	X	Y	X	Y
abbey	0.2639	0.2978	0.2824	0.2805	0.874	0.2758	0.2896	0.2739
dublin	0.0382	0.0431	0.0406	0.0404	0.0411	0.0395	0.0408	0.0387
fifth	0.0383	0.0435	0.0409	0.0408	0.0412	0.0398	0.0409	0.0390
italy	0.2584	0.2923	0.2764	0.2752	0.2810	0.2705	0.2828	0.2684
rio	0.0378	0.0429	0.0403	0.0402	0.0407	0.0393	0.0404	0.0385
shibuya	0.2938	0.3341	0.3143	0.3151	0.3188	0.3095	0.3206	0.3070
times	0.2664	0.3015	0.2849	0.2840	0.2897	0.2791	0.2918	0.2771
times2	0.2670	0.3021	0.2857	0.2845	0.2905	0.2797	0.2926	0.2776
times3	0.0385	0.0438	0.0411	0.0410	0.0415	0.0401	0.0411	0.0392
wallst	0.2604	0.2930	0.2788	0.2757	0.2841	0.2713	0.2866	0.2695

Tabela 4.4: Coordenadas  $(x, y)$  para os vídeos agitados com extração retangular.

Nome do Vídeo	Três Cortes		Cinco Cortes		Sete Cortes		Onze Cortes	
	X	Y	X	Y	X	Y	X	Y
abbey	0.1975	0.2243	0.2118	0.2121	0.2143	0.2076	0.2161	0.2066
dublin	0.0270	0.0678	0.0292	0.0292	0.0294	0.0285	0.0292	0.0279
fifth	0.0154	0.0174	0.0164	0.0163	0.0165	0.0159	0.0162	0.0154
italy	0.0413	0.0466	0.0442	0.0438	0.0448	0.0429	0.0451	0.0426
rio	0.0276	0.0313	0.0294	0.0294	0.0296	0.0286	0.0294	0.0281
shibuya	0.4492	0.5092	0.4809	0.4806	0.4888	0.4725	0.4916	0.4690
times	0.4267	0.4821	0.4569	0.4550	0.0638	0.0616	0.4685	0.4448
times2	0.0590	0.0669	0.0631	0.0630	0.2905	0.2797	0.0642	0.0611
times3	0.0803	0.0911	0.0855	0.0855	0.0870	0.0841	0.0871	0.0831
wallst	0.0271	0.0306	0.0288	0.0288	0.0290	0.0280	0.0289	0.0275

Tabela 4.5: Coordenadas  $(x, y)$  para os vídeos calmos com extração por arcos retangulares.

Nome do Vídeo	Três Cortes		Cinco Cortes		Sete Cortes		Onze Cortes	
	X	Y	X	Y	X	Y	X	Y
axe	0.3235	0.2313	0.3394	0.2072	0.3418	0.2036	0.3423	0.2030
crowdwave	0.0468	0.0337	0.0485	0.0309	0.0488	0.0308	0.0486	0.0323
fight	0.0472	0.0339	0.0490	0.0311	0.0492	0.0310	0.0490	0.0325
fight4	0.3178	0.2270	0.3329	0.2032	0.3351	0.1996	0.3354	0.1988
fight5	0.0464	0.0335	0.0483	0.0308	0.0485	0.0308	0.0484	0.0323
fight7	0.3628	0.2573	0.3810	0.2313	0.3828	0.2268	0.3826	0.2257
kiev	0.3276	0.2338	0.3438	0.2097	0.3460	0.2059	0.3463	0.2052
police	0.3282	0.2344	0.3442	0.2101	0.3464	0.2063	0.3468	0.2056
pub	0.0475	0.0341	0.0493	0.0313	0.0495	0.0312	0.0493	0.0327
show	0.3183	0.2284	0.3335	0.2043	0.3363	0.2010	0.3372	0.2005

Tabela 4.6: Coordenadas  $(x, y)$  para os vídeos agitados com extração por arcos retangulares.

Nome do Vídeo	Três Cortes		Cinco Cortes		Sete Cortes		Onze Cortes	
	X	Y	X	Y	X	Y	X	Y
axe	0.2444	0.1737	0.2563	0.1558	0.2578	0.1530	0.2585	0.1528
crowdwave	0.0339	0.0244	0.0352	0.0224	0.0354	0.0223	0.0353	0.0235
fight	0.0191	0.0137	0.0198	0.0124	0.0197	0.0127	0.0194	0.0145
fight4	0.0507	0.0367	0.0530	0.0329	0.0529	0.0333	0.0534	0.0335
fight5	0.0341	0.0244	0.0353	0.0224	0.0355	0.0223	0.0355	0.0235
fight7	0.5549	0.3929	0.5819	0.3520	0.5873	0.3427	0.5904	0.3360
kiev	0.5250	0.3737	0.5508	0.3343	0.5561	0.3255	0.5597	0.3193
police	0.0731	0.0523	0.0765	0.0469	0.0767	0.0467	0.0768	0.0479
pub	0.0992	0.0708	0.1036	0.0637	0.1041	0.0630	0.1041	0.0646
show	0.0334	0.0240	0.0346	0.0220	0.0348	0.0220	0.0348	0.0231

# Capítulo 5

## Conclusão e Trabalhos Futuros

Esse projeto propôs um meio de desenvolver nova técnica para analisar eventos em ambientes monitorados por câmeras de segurança. A hipótese inicial do problema, de que as altas frequências do campo de movimento estimado do fluxo ótico podem detectar variações no comportamento de uma cena, pode ser comprovada pelos resultados apresentados no Capítulo 4, pois há evidências de que é possível construir um classificador estável e confiável para o problema relatado.

O próximo passo é verificar a correlação entre número de cortes e novas maneiras de extração de informações, fazendo testes com valores maiores do que os apresentados e procurando novas formas de corte, como uma formatação em blocos nas cantos superiores e inferiores da componente de alta frequência. Então, desenvolver um classificador robusto, como o de rede neural, será essencial para concluir um sistema de detecção de eventos em multidões completo.

Uma base de dados mais vasta e complexa também é necessário, visto que a base de dados original é pequena e os vídeos possuem curta duração.

A otimização das técnicas usadas para a proposta do descritor também são um projeto futuro. Os testes com outras técnicas de fluxo ótico, como o método diferencial de Lucas-Kanade[20], serão realizados para verificar se houve melhora nos resultados obtidos. A Transformada de Fourier também será modificada para uma Transformada de Cossenos, afim de observação de resultados. Com essa Transformada de Cossenos, é possível criar mais uma técnica de extração das informações, com um corte diagonal, visto que a Transformada divide suas frequências nas diagonais superiores e inferiores.

O desenvolvimento de um classificador robusto como *Support Vector Machine*[4] e redes neurais artificiais também fazem parte de uma extensão do projeto. Além disso, é importante incluir a implementação da estratégia proposta em uma linguagem de programação de alto nível para que as operações possam ser realizadas em tempo real, incluindo

análise de tempo, usando vídeos reais e também fazer comparações com as últimas técnicas propostas para detecção e reconhecimento de eventos em multidões.

# Referências

- [1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, 1989. 7
- [2] J.L. Barron, D.J. Fleet, S.S. Beauchemin, and T.A. Burkitt. Performance of optical flow techniques. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pages 236–242, Jun 1992. 7
- [3] G. Bradski. *Dr. Dobb's Journal of Software Tools*. 4, 13
- [4] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 43
- [5] Flávio de Barros Vidal. Estudo do movimento em imagens através do fluxo óptico determinado pelos métodos diferencial, de similaridade de regiões e de filtragem espaço-temporal. Master's thesis, Universidade de Brasília, 2005. x, 10
- [6] E. Esen, M.A. Arabaci, and M. Soysal. Fight detection in surveillance videos. In *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, pages 131–135, June 2013. 4
- [7] Claude L Fennema and William B Thompson. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9(4):301 – 315, 1979. 8
- [8] David J. Fleet and Allan D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, pages 77–104, 1990. 7
- [9] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Always learning. Pearson, 2012. 5
- [10] C. Garate, P. Bilinsky, and F. Bremond. Crowd event recognition using hog tracker. In *Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009 Twelfth IEEE International Workshop on*, pages 1–6, Dec 2009. 3
- [11] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2 edition, 2002. x, 11, 12, 13
- [12] David J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, pages 279–302, 1988. 7

- [13] B. K. P. Horn and B. G. Schunck. Determining optical flow. In *Artificial Intelligence*, pages 185–204, 1981. 8, 10, 15, 17
- [14] M. Husni and N. Suryana. Crowd event detection in computer vision. In *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, volume 1, pages V1–444–V1–447, July 2010. 1
- [15] Yan Ke, R. Sukthankar, and M. Hebert. Event detection in crowded videos. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007. 1, 3
- [16] H. Kruegle. *CCTV Surveillance: Video Practices and Technology*. CCTV Surveillance Series. Elsevier Science, 2011. 15
- [17] R. Laganière. *OpenCV 2 Computer Vision Application Programming Cookbook*. Packt open source. Lightning Source Incorporated, 2011. 4
- [18] Guohui Li, Jun Chen, Boliang Sun, and Haozhe Liang. Crowd event detection based on motion vector intersection points. In *Computer Science and Information Processing (CSIP), 2012 International Conference on*, pages 411–415, Aug 2012. 3
- [19] Honghong Liao, Jinhai Xiang, Weiping Sun, Qing Feng, and Jianghua Dai. An abnormal event recognition in crowd scene. In *Image and Graphics (ICIG), 2011 Sixth International Conference on*, pages 731–736, Aug 2011. 3
- [20] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. pages 674–679, 1981. 43
- [21] Eadweard Muybridge. *The horse in motion*, 1878. x, 4
- [22] Dong Wang, Zhang Zhang, Wei Wang, Liang Wang, and Tieniu Tan. Baseline results for violence detection in still images. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pages 54–57, Sept 2012. 4
- [23] A.M. Waxman, J. Wu, and Fredrik Bergholm. Convected activation profiles and the measurement of visual motion. In *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*, pages 717–723, Jun 1988. 8
- [24] Long Xu, Chen Gong, Jie Yang, Qiang Wu, and Lixiu Yao. Violent video detection based on mosift feature and sparse coding. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3538–3542, May 2014. 4